Semiparametric Methods for Causal Mediation Analysis and Measurement Error

Citation	Miles, Caleb Hilliard. 2015. Semiparametric Methods for Causal Mediation Analysis and Measurement Error. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences.
Citable link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:23845420
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Semiparametric Methods for Causal Mediation Analysis and Measurement Error

A dissertation presented

by

Caleb Hilliard Miles

to

The Department of Biostatistics

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Biostatistics

Harvard University
Cambridge, Massachusetts

July 2015

Dissertation Advisor: Professor Eric J. Tchetgen Tchetgen Caleb Hilliard Miles

Semiparametric Methods for Causal Mediation Analysis and Measurement Error

Abstract

Chapter 1: Since the early 2000s, evidence has accumulated for a significant differential effect of first-line antiretroviral therapy (ART) regimens on human immunodeficiency virus (HIV) treatment outcomes, such as CD4 response and viral load suppression. This finding was replicated in our data from the Harvard President's Emergency Plan for AIDS Relief (PEPFAR) program in Nigeria. Investigators were interested in finding the source of these differences, i.e., understanding the mechanisms through which one regimen outperforms another, particularly via adherence. This amounts to a mediation question with adherence playing the role of mediator. Existing mediation analysis results, however, have relied on an assumption of no exposure-induced confounding of the intermediate variable, and generally require an assumption of no unmeasured confounding for nonparametric identification. Both assumptions are violated by the presence of drug toxicity. In this paper, we relax these assumptions and show that certain path-specific effects remain identified under weaker conditions. We focus on the path-specific effect solely mediated by adherence and not by toxicity and propose a suite of estimators for this effect, including a semiparametric-efficient, multiply-robust estimator. We illustrate with simulations and present results from a study applying the methodology to the Harvard PEPFAR data.

Chapter 2: In causal mediation analysis, nonparametric identification of the pure (natural) direct effect typically relies on fundamental assumptions of (i) so-called "cross-world-counterfactuals" independence and (ii) no exposure-induced confounding. When the mediator is binary, bounds for partial identification have been given when neither assumption is made, or alternatively when assuming only (ii). We extend these bounds to the case of a polytomous mediator, and provide bounds for the case assuming only (i). We apply these bounds to data from the Harvard PEPFAR program in Nigeria, where we evaluate the extent to which the effects of antiretroviral therapy on virological failure are mediated by a patient's adherence, and show that inference on this effect is somewhat sensitive to model assumptions.

Chapter 3: When assessing the presence of an exposure causal effect on a given outcome, it is well known that classical measurement error of the exposure can seriously reduce the power of a test of the null hypothesis in question, although its type I error rate will generally remain controlled at the nominal level. In contrast, classical measurement error of a confounder can have disastrous consequences on the type I error rate of a test of treatment effect. In this paper, we develop a large class of semiparametric test statistics of an exposure causal effect, which are completely robust to classical measurement error of a subset of confounders. A unique and appealing feature of our proposed methods is that they require no external information such as validation data or replicates of error-prone confounders. The approach relies on the observation that under the sharp null hypothesis of no exposure causal effect, the standard assumption of no unmeasured confounding implies that the outcome is in fact a valid instrumental variable for the association between the error-prone confounder and the exposure. We present a doubly-robust form of this test that requires only one of two models – an outcome-regression and a propensity-score model – to be correctly specified for the resulting test statistic to have correct type I error rate. Validity and power within our class of test statistics is demonstrated via multiple simulation studies. We apply the methods to a multi-U.S.-city, time-series data set to test for an effect of temperature on mortality while adjusting for atmospheric particulate matter with diameter of 2.5 micrometres or less (PM2.5), which is well known to be measured with error.

Contents

Title page i

Copyright page ii

Abstract iii

Table of Contents v

Contents **v**

Acknowledgments vii

1 Quantifying an Adherence Path-Specific Effect of Antiretroviral Therapy in the Nigeria PEPFAR Program **1**

1.1 Introduction 2

1.2 Definitions 6

1.3 Identification 9

1.4 Path-specific inference 10

1.4.1 Maximum Likelihood Estimation 11

1.4.2 Multiply-Robust Estimation 11

1.4.3 Other Estimators 13

1.5 Simulation study 14

1.6 Harvard PEPFAR Nigeria analysis 17

1.7 Discussion 21

2 On Partial Identification of the Pure Direct Effect **24**

2.1 Introduction 25

2.2 Preliminaries 27

2.3 New partial identification results 33

2.4 Application to Harvard PEPFAR data set 37

3 A Class of Semiparametric Tests of Treatment Effect Robust to Measurement Error of a Confounder 42

3.1 Introduction 43

3.2 A class of propensity-score-based test statistics robust to measurement error 46

3.3 A simulation study of the test statistic $\chi^2_{\text{robust}_A}$ 49

3.4 A class of doubly-robust test statistics 50

3.5 A simulation study for double robustness 52

3.6 Application to test for an effect of temperature on mortality 53

3.7 Power and estimation under an additive causal model 56

3.8 Extensions to binary and count exposures 58

3.9 Discussion 59

References 61

A Proofs and theoretical results for Chapter 1 72

A.1 Identification Result Proof 72

A.2 Derivation of Estimation Strategies 73

A.2.1 Maximum Likelihood Estimator 73

A.2.2 Estimator a 73

A.2.3 Estimator b 74

A.3 Derivation of the Influence Function 74

A.4 Multiple-Robustness of the Efficient Influence Function 79

B Additional stabilization technique for the multiply-robust estimator 82

C Plots comparing estimators in the PEPFAR Nigeria study 83

D Proofs for Chapter 2 88

E Proofs and additional figures for Chapter 3 90

Acknowledgments

I owe a deep debt of gratitude to my adviser, Eric Tchetgen Tchetgen, for not only his phenomenal academic guidance and helping shape my view of statistics, but also his patience, support, and understanding during times in which it was most needed. I feel truly honored and privileged to have had the opportunity to work with him.

I sincerely thank my dissertation committee for their many contributions of scientific expertise. Additionally, I thank Brent Coull for his ever-encouraging voice, Phyllis Kanki for her warm and welcoming spirit, and Tyler Vanderweele for his generosity with his time and wisdom. I would also like to thank my other coauthors, Ilya Shpitser, Seema Meloni, and Joel Schwartz, for their excellent insight and enthusiasm.

I am indebted to the wonderful friends I have been lucky enough to have made during my time in the Boston area. They have provided an essential support system and have made my time here unforgettable. In particular, my deepest thanks go to Matt and Megan Fisher, Sarah Matousek, and Luis Gonzalez for all of their love and support. I am also very thankful to my cohort for creating a nurturing environment and for their eagerness to help with any problem. In particular, I thank Yered Pita-Juarez and Matey Neykov for their friendship and the many ways in which they have assisted me during my time as a doctoral student.

A tremendous thanks goes to my family. I would like to thank my parents, Caleb and Cathy Miles, and grandparents (especially my sweet Grandmama who we lost in 2013) for always loving and supporting me, for prioritizing my education, and for having an often unrealistic level of confidence in me. I would like to thank my brother, Taylor, for his kind heart and gentle spirit, and for inspiring me to be a better person.

Lastly, I thank God for his many unearned blessings and opportunities, without which I would not be where I am today.

Quantifying an Adherence Path-Specific Effect of Antiretroviral Therapy in the Nigeria PEPFAR Program

Caleb H. Miles, Ilya Shpitser, Phyllis Kanki, Seema Meloni, and Eric J. Tchetgen Tchetgen

Department of Biostatistics

Harvard School of Public Health

1.1 Introduction

The President's Emergency Plan for AIDS Relief (PEPFAR) has been a highly successful program that has saved millions of lives worldwide since its inception in 2003. The Harvard School of Public Health was awarded one of the PEPFAR grants, receiving a total of $362 million for work in Nigeria, Botswana, and Tanzania. The program has furnished these countries with invaluable medical infrastructure and supported AIDS care services in Nigeria for over 160,000 people and treatment to approximately 105,000 of those patients.

Our data set consists of previously antiretroviral therapy (ART)-naïve, human immunodeficiency virus (HIV)-1 infected, adult patients enrolled in the Harvard PEPFAR/AIDS Prevention Initiative in Nigeria (APIN) program between June 2004 and November 2010 who started ART in the program and were followed for at least 1 year after initiating ART. Upon entry into the Harvard/APIN PEPFAR HIV care program, all patients completed informed consent; all consent forms were approved by the institutional review boards at Harvard, APIN and all the corresponding Harvard/APIN PEPFAR HIV care and treatment sites. Patients not on one of 6 standard first-line regimens at baseline or seen at two of the hospitals without reliable viral load data were excluded from the data set. The analysis in this paper consists of only the complete cases, and results are given for all regimens but d4T+3TC+EFV (see Table 1.1 note for full drug names), due to the small sample of patients on this regimen as a consequence of it having been dropped midway through the program. d4T+3TC+NVP was also dropped, but had a large enough sample to provide for stable inference.

The significant funding support for AIDS treatment in resource-limited settings provided by PEPFAR and other international donor organizations relied on clinical trial data generated in resource rich settings. In order to maximize the benefit of providing ART to the largest number of patients, well-established drug regimens that were less costly were recommended and supported by the program. Studies dating back to the early 2000s have demonstrated evidence that these first-line regimens were not equally effective (Tang et al., 2012), and indeed, in the Harvard PEPFAR data, we have observed a significant differential effect of first-line ART regimens on virologic failure and, to a lesser extent, CD4 count. Since these were first-line regimens in use in most resource-limited settings, this difference could have widespread implications to the success of ART pro-

Table 1.1: Treatment regimen coding and their estimated average causal effects on risk of virologic failure (VF) and CD4 count

		Patients on	RR of	log-RR of	Mean diff. in
Code	ART regimen	regimen	VF (s.e.)	VF (s.e.)	CD4 count
1	TDF + 3TC/FTC + EFV	1448 (14.6%)	0.65 (0.014)	-0.44 (0.12)	6.9 (7.4)
2	d4T + 3TC + NVP	854 (8.6%)	0.75 (0.017)	-0.29 (0.13)	-9.7 (6.8)
3	AZT + 3TC + EFV	1003 (10.1%)	0.78 (0.018)	-0.25 (0.14)	10.7 (8.7)
4	AZT + 3TC + NVP	4707 (47.4%)	0.82 (0.011)	-0.21 (0.078)	17.8 (4.9)
5	TDF + 3TC/FTC + NVP	1919 (19.3%)	-	-	-

NOTE: 3TC=lamivudine, AZT=zidovudine, d4T=stavudine, EFV=efavirenz, FTC=emtricitabine, NVP=nevirapine, TDF=tenofovir. Effects on risk of virologic failure are expressed on the risk ratio (RR) and log-risk ratio scale relative to treatment 5 and were estimated using inverse-probability weighted estimators. Effects on CD4 count are expressed on the mean difference scale relative to treatment 5 and were estimated using doubly-robust estimators. All effects adjusted for the confounders listed in Section 1.2.

grams. These regimens and each of their corresponding total effects on virologic failure and CD4 count relative to a common reference treatment are reported in Table 1.1. The effects on virologic failure are reported as marginal and log-marginal risk ratios, and the effects on CD4 count and log CD4 count are reported on the mean-difference scale. Treatments were coded from strongest estimated effect on virologic failure to weakest, and the weakest treatment (TDF+3TC/FTC+NVP) was chosen as the reference for the purposes of the effects in Table 1.1. These effects are contrasts in the population between the risk of virologic failure had one intervened to assign everyone to a comparison-level treatment (1, 2, 3, or 4) and that if one had intervened to assign everyone to baseline treatment 5. The total effects of these regimens, however, do not quite tell the whole story. Investigators were interested in finding the source of these differences, i.e., understanding the mechanisms through which one regimen outperforms another. Mediation analysis serves to better explain these mechanisms that drive the differences in effects. This type of analysis has the potential to help target interventions to improve the performance of the less-robust regimens.

The total effect can be considered as a combination of effects, possibly in conflicting directions, through different pathways from the exposure to the outcome. Therefore, a weak total effect could be due to a combination of even weaker path-specific effects or several stronger path-specific effects canceling one another out. One such path-specific effect could work strictly through biological pathways, in which case this population would benefit most from switching to a more favorable drug regimen. Alternatively, biological factors might play a comparatively smaller role relative to the effect of the treatment through nonbiological pathways, such as through adherence (Shpitser, 2013). We suspect a lack of adherence to treatment to be a driving mechanism of the observed differential effects, in which case it would be worth considering how to improve this mediating factor.

Adherence is widely accepted as a key factor for sustained viral suppression and is considered a prerequisite for maintenance on a prescribed drug regimen and optimal patient outcomes. However, the extent to which adherence to a given choice of first-line ART contributes to virologic failure (defined by the World Health Organization [WHO] as repeat viral load > 1000 copies/mL after 6 months of ART duration) is complex and still poorly understood and is a pressing mediation question in HIV research (Bangsberg et al., 2000). Understanding this issue is particularly important in resource-limited settings, where ART regimen options are few, and adherence to lifelong multi-drug daily dosing is challenging, but necessary. In such settings, quantifying to what degree differential rates of virologic failure are due to differences in adherence rates between therapies would inform the extent to which failure rates could be reduced by programs that improve adherence rates for certain ARTs, rather than changing the ART regimens themselves. Such adherence interventions have been very successful in the treatment of tuberculosis (China Tuberculosis Control Collaboration, 1996; Fujiwara et al., 1997; Suárez et al., 2001) and are considered similarly important in the treatment of HIV (Mills et al., 2006; Vranceanu et al., 2008; Pop-Eleches et al., 2011).

Remark. *Technically, the WHO also requires demonstration of adherence in their definition of virologic failure, which we avoid using in this paper since we cannot study the role of adherence as a mediator when it is part of the definition of the outcome.*

Among other potential mechanisms, the effect of treatment on virologic failure and CD4 count may be mediated by adherence, drug toxicity, or both. This study investigates the extent to which adherence, and not drug toxicity, mediates the effect, using the Harvard PEPFAR data set. That is, we focus on the role of adherence when it is differentially affected by the ways the drugs are obtained and taken, rather than by different levels of toxic side effects. The effect mediated by nonadherence due to toxicity is unlikely to be appreciable, since toxicity in Nigeria is typically clinically recognizable and actionable. The magnitude of the roles of other drug-specific predictors of nonadherence, on the other hand, are less understood. These predictors also potentially point to lower-hanging fruit for development of adherence-promoting interventions. In mediation analysis terminology, we aim to estimate the effects of treatment assignment on virologic failure and CD4 count that are indirect with respect to adherence but direct with respect to toxicity. The definition, identification, and estimation of direct and indirect effects have received much attention in recent causal inference literature (Robins and Greenland, 1992; Robins, 1999, 2003; Pearl, 2001; Avin et al., 2005; Taylor et al., 2005; Petersen et al., 2006; Ten Have et al., 2007; Goetgeluk et al., 2008; van der Laan and Petersen, 2008; VanderWeele, 2009, 2011; VanderWeele and Vansteelandt, 2009, 2010; Imai et al., 2010a,b; Tchetgen Tchetgen, 2011; Tchetgen Tchetgen and Shpitser, 2014, 2012; Tchetgen Tchetgen, 2013).

The particular effect we are interested in can be classified as a *path-specific effect* (Pearl, 2001) – a class of estimands which can represent effects along any given causal pathway or collection of causal pathways. We consider the effect along the path from the provision of ART to virologic failure (or CD4 count) that goes through adherence, but not through toxicity. This effect is a measure of the change in risk of virologic failure (or mean CD4 count) were one to intervene on the mechanism by which the choice of treatment regimen directly, i.e., not through toxicity, affects adherence. For instance, if the difference in the effectiveness of ART through adherence were due to some regimens of ART having certain meal restrictions, posing a greater risk of patients missing dosages due to issues with food insecurity (Eldred et al., 1998; Gifford et al., 1998; Roberts, 2000), this effect would reflect the change in mean outcome if we were to modify the pills such that they can be taken without any meal restrictions. We emphasize our focus on the pathway through adherence, which does not involve toxicity, to learn about other possible mediating mechanisms that may be as important as toxicity, but are currently underappreciated. The presence of an effect through this pathway calls for closer investigation of these possible mechanisms, such as number of pills taken per dosage or the requirement that they be taken with meals. The absence indicates that differential effects through other pathways are driving the observed differences in effects among the treatment assignments. In particular, the efficacies of the drugs themselves may, in fact, differ, i.e., they may have a differential direct effect on the outcome with respect to adherence, or they may have a differential effect on adherence due to their differing levels of toxicity.

Pearl (2001) defines path-specific effects, and Avin et al. (2005) provide general necessary and sufficient conditions for their identification for a single exposure and outcome, while Shpitser (2013) generalizes these definitions and conditions to settings with multiple exposures, multiple outcomes, and possible hidden variables. Our path-specific effect described above satisfies these identifying conditions, however an estimation strategy for its identifying functional does not yet exist. In this paper, we develop a suite of estimators (including a multiply-robust, semiparametric-efficient estimator) for the effect. The HIV case study detailed in this paper also functions as a guide for the application of this new method to analogous mediation settings where there is confounding that is affected by the exposure.

1.2 Definitions

To formalize our discussion, we begin by defining variables and counterfactuals. We will be considering pairwise comparisons of first-line ARTs prescribed to most HIV patients in Nigeria. Let E be an indicator of exposure to one of two such regimens of ART (coding given in Table 1.1). For notational simplicity, let e' denote the “reference level" treatment and e denote the “comparison level" treatment. Let $\mathbf{C_1}$ be a bivariate vector of an indicator of any lab toxicities (alanine transaminase $\geq$120 UI/L, Creatinine $\geq$260 mmol/L, Hemoglobin $\leq$8 g/dL) observed six months after treatment initiation and an indicator that the patient’s average percent adherence during the same six months, i.e., the total number of days that the patient had their drug supply divided by the number of days in the six month period, was no less than 95%. Let M be an indicator that the patient’s average percent adherence during the subsequent six months was no less than 95%. Let Y be an indicator of whether the patient experienced virologic failure at the end of the year (based on viral load measurements at twelve and eighteen months for confirmation), or alternatively CD4 count at twelve months. Let $\mathbf{C_0}$ be a vector of baseline confounders of the causal relationships between E, M, and Y not affected by exposure, viz. sex, age, marital status, WHO stage, hepatitis C virus, hepatitis B virus, CD4 count, and viral load. Throughout, we will assume that we observe i.i.d. sampling of $\mathbf{O} = (\mathbf{C_0}, E, \mathbf{C_1}, M, Y)$.

We now consider counterfactuals under possible interventions on the variables (Rubin, 1974,

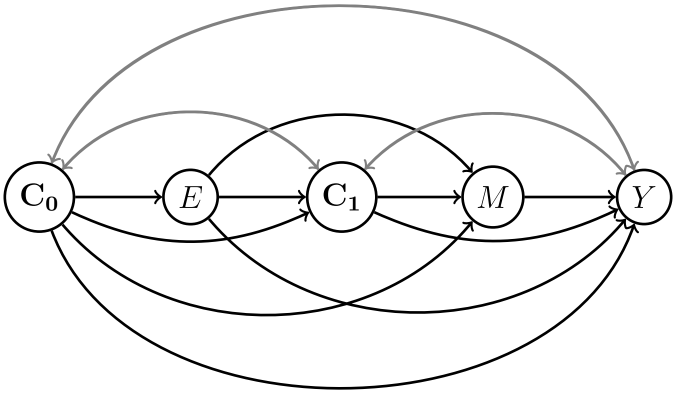


Figure 1.1: A causal graph with unobserved confounders that allows for identification of the $\mathcal{P}_{EMY}$-specific effect

1978). Let $Y(e^*)$ denote a patient's virologic suppression status or CD4 count if assigned, possibly contrary to fact, to the regimen of ART e^*. In the context of mediation, there will also be counterfactuals for intermediate variables. We define $\mathbf{C_1}(e^*)$, $M(\mathbf{c_1}, e^*)$ and $Y(m, e^*)$ similarly, and adopt the standard set of consistency assumptions (Robins, 1986) that if $E = e^*$, then $\mathbf{C_1}(e^*) = \mathbf{C_1}$ w.p.1, if $E = e^*$ and $\mathbf{C_1} = \mathbf{c_1}$, then $M(\mathbf{c_1}, e^*) = M$ w.p.1, if $E = e^*$ and $M = m$, then $Y(m, e^*) = Y$ w.p.1, and if $E = e^*$, then $Y(e^*) = Y$ w.p.1. Additionally, we adopt the standard set of positivity assumptions (Robins, 1986) that $f_{M|\mathbf{C_1},E,\mathbf{C_0}}(m|\mathbf{C_1}, E, \mathbf{C_0}) > 0$ w.p.1 for each $m \in \text{supp}(M)$, $f_{\mathbf{C_1}|E,\mathbf{C_0}}(\mathbf{c_1}|E, \mathbf{C_0}) > 0$ w.p.1 for each $\mathbf{c_1} \in \text{supp}(\mathbf{C_1})$, $f_{E|\mathbf{C_0}}(e^*|\mathbf{C_0}) > 0$ w.p.1 for each $e^* \in \{e', e\}$.

To define the path-specific effect along the path $E \to M \to Y$, which we denote $\mathcal{P}_{EMY}$, we begin by discussing the graph in Figure 1.1. This is a complete graph of all observed variables in the sense that it includes all possible directed arrows that follow the natural temporal ordering. That is, any variable may directly affect any other variable succeeding it under this graph. The graph departs from the standard mediation graph (Baron and Kenny, 1986) in two important ways.

The first is with the presence of $\mathbf{C_1}$, which allows confounders of the effect of the mediator on the outcome to be affected by the exposure. In our HIV context, $\mathbf{C_1}$ contains toxicity, which is clearly affected by the treatment assignment and may confound the effect of adherence on virologic failure. One way in which it might do this is on a biological level, toxicity might have an interactive effect with the drugs on the outcome, allowed for by the presence of the directed arrow from $\mathbf{C_1}$ to Y in conjunction with the directed arrow from E to Y. Thus, toxicity is a common cause of the outcome and adherence and, therefore, a confounder. Such a confounder is known as a *recanting witness*, due to its role in telling two conflicting “stories" about how E affects Y by being involved in two different pathway from E to Y – one involving M and the other not. Avin et al. (2005) showed the natural (or pure) direct and indirect effects (NDE and NIE, both highly popular in the mediation literature) (Robins and Greenland, 1992; Pearl, 2001) to be unidentified in the presence of a recanting witness.

The second way the graph in Figure 1.1 departs from the standard mediation graph is by the presence of the gray bidirected edges between $\mathbf{C_0}$, $\mathbf{C_1}$, and Y, each of which represents unobserved common causes between the two nodes to which it points. In the HIV application, these bidirected edges allow for the possibility of underlying biological factors which may be unobserved common causes of toxicity, the outcome, and biological baseline measurements such as viral load. The presence of these bidirected edges induces confounding of the effect of adherence on the outcome via toxicity, even if the arrow directed from $\mathbf{C_1}$ to Y is absent. Since early adherence (during the first six months) may confound the effect of adherence at a later stage (during the subsequent six months) on the outcome, early adherence must be included in $\mathbf{C_1}$. Thus, $\mathcal{P}_{EMY}$ involves only later adherence, and neither toxicity nor early-stage adherence.

As described above, we wish to quantify the mediating role of adherence along $\mathcal{P}_{EMY}$ in Figure 1.1 which does not involve toxicity. Effects along such arbitrary (bundles of) causal pathways are known as path-specific effects (Pearl, 2000; Avin et al., 2005; Shpitser, 2013) and it is possible to define them inductively, which results in a quantity that is a function of a nested counterfactual (Shpitser, 2013). A general definition for the static-treatment and single-outcome case is given by Pearl (2000) and Avin et al. (2005). Defining

$$\beta_0 \equiv \mathbb{E}[Y(M(e, \mathbf{C_1}(e')), \mathbf{C_1}(e'), e')]$$
$$\delta_0 \equiv \mathbb{E}[Y(M(e', \mathbf{C_1}(e')), \mathbf{C_1}(e'), e')],$$

the $\mathcal{P}_{EMY}$-specific effect, with respect to the comparison treatment value e and the baseline treatment value e' on the mean difference scale, is given by $\beta_0 - \delta_0$. δ_0 gives the mean outcome had everyone been assigned to the reference treatment regimen. β_0 gives the mean outcome had everyone been assigned to the reference treatment regimen, and adhered as they would have based on the toxicity they experienced from this regimen, but otherwise as if they had been assigned to the comparison treatment. This is the $\mathcal{P}_{EMY}$-specific effect since it captures the impact of changing $M(e')$ to $M(e, \mathbf{C_1}(e'))$, which in turn would lead to an effect on Y only if M affects Y directly when all patients are assigned to e'.

1.3 Identification

Before introducing our identification result, we must first introduce a model that relaxes the assumption of independent errors of the Markovian model (Pearl, 2000) in a natural way. We will associate this model with the graph in Figure 1.1. This model consists of a set of equations, one for each variable in the graph. With each random variable on the graph is associated a distinct, arbitrary function, denoted g, and a distinct random disturbance, denoted ε, each with a subscript corresponding to its respective random variable. A component in a graph connected by bidirected edges (i.e., connected when ignoring directed edges) is known as a *district* (Richardson, 2009) or *c-component* (Tian and Pearl, 2002). The sets of random disturbances corresponding to each district are assumed to be mutually independent of one another. That is, $\{\boldsymbol{\varepsilon}_{\mathbf{C_0}}, \boldsymbol{\varepsilon}_{\mathbf{C}_1}, \varepsilon_Y\}$, ε_E, and ε_M are mutually independent; $\boldsymbol{\varepsilon}_{\mathbf{C_0}}$, $\boldsymbol{\varepsilon}_{\mathbf{C_1}}$, and ε_Y, however, are not. Each variable is generated by its corresponding function, which depends only on all variables that directly affect it (i.e., its parents on the graph), and its corresponding random disturbance, as follows: $\mathbf{C_0} = \mathbf{g}_{\mathbf{C_0}}(\boldsymbol{\varepsilon}_{\mathbf{C_0}})$, $E = g_E(\mathbf{C_0}, \varepsilon_E)$, $\mathbf{C_1} = \mathbf{g}_{\mathbf{C_1}}(\mathbf{C_0}, E, \boldsymbol{\varepsilon}_{\mathbf{C_1}})$, $M = g_M(\mathbf{C_0}, E, \mathbf{C_1}, \varepsilon_M)$, $Y = g_Y(\mathbf{C_0}, E, \mathbf{C_1}, M, \varepsilon_Y)$.

Just as the Markovian model, the model we introduce is especially useful for making counterfactual independence assumptions explicit. Take for instance the statement $\{Y(m, e'), \mathbf{C_1}(e')\} \perp\!\!\!\perp M(c_1, e) | \mathbf{C_0}$. To see whether this statement holds in the context of the graph in Figure 1.1, observe what occurs when we intervene on the mechanism in one case to force the exposure to be the comparison level, e, and set $\mathbf{C_1}$ to an arbitrary value $\mathbf{c_1}$: $\mathbf{C_0} = \mathbf{g}_{\mathbf{C_0}}(\boldsymbol{\varepsilon}_{\mathbf{C_0}})$, $E = e$, $\mathbf{C_1} = \mathbf{c_1}$, $M(\mathbf{c_1}, e) = g_M(\mathbf{C_0}, e, \mathbf{c_1}, \varepsilon_M)$, $Y(\mathbf{c_1}, e) = g_Y(\mathbf{C_0}, e, \mathbf{c_1}, M(\mathbf{c_1}, e), \varepsilon_Y)$; and in another case to force the exposure to be the reference level, e', and set M to an arbitrary value m: $\mathbf{C_0} = \mathbf{g}_{\mathbf{C_0}}(\boldsymbol{\varepsilon}_{\mathbf{C_0}})$, $E = e'$, $\mathbf{C_1}(e') = \mathbf{g}_{\mathbf{C_1}}(\mathbf{C_0}, e', \boldsymbol{\varepsilon}_{\mathbf{C_1}})$, $M = m$, $Y(m, e') = g_Y(\mathbf{C_0}, e', \mathbf{C_1}(e'), m, \varepsilon_Y)$. Note that the only sources of stochasticity in $M(\mathbf{c_1}, e)$ are $\mathbf{C_0}$ and ε_M, and the only sources of stochasticity in

$\{Y(m, e'), \mathbf{C_1}(e')\}$ are $\mathbf{C_0}$, $\boldsymbol{\varepsilon}_{\mathbf{C_1}}$, and ε_Y. Hence the only source of dependence between the two is $\mathbf{C_0}$ since $\varepsilon_M \perp\!\!\!\perp \{\boldsymbol{\varepsilon}_{\mathbf{C_1}}, \varepsilon_Y\}$, and they are independent conditional on $\mathbf{C_0}$. We are now prepared to present our identification result, whose proof is provided in the supplementary materials.

Theorem 1.1. *Suppose the data-generating mechanism from which the observed data* $\mathbf{O}$ *are sampled follows the relaxation of the Markovian model that we introduce above, represented by the graph in Figure 1.1. Then* β_0 *is identified under this model by the following functional of* $F_{\mathbf{O}}$*:*

$$\beta_0 = \iiint\limits_{m,\mathbf{c_1},\mathbf{c_0}} \mathbb{E}(Y|m, \mathbf{c_1}, e', \mathbf{c_0}) dF(m|\mathbf{c_1}, e, \mathbf{c_0}) dF(\mathbf{c_1}|e', \mathbf{c_0}) dF(\mathbf{c_0}). \tag{1.1}$$

Remark. *The following conditions are sufficient for the same identification result, and are strictly weaker than those implied by our model: for all* m*,* $\mathbf{c_1}$*,* e*, and* e'*,* $\{Y(m, e'), \mathbf{C_1}(e')\} \perp\!\!\!\perp E|\mathbf{C_0}$*,* $Y(m) \perp\!\!\!\perp M|\mathbf{C_1}, E, \mathbf{C_0}$*,* $M(\mathbf{c_1}, e) \perp\!\!\!\perp \{\mathbf{C_1}, E\}|\mathbf{C_0}$*,* $\{Y(m, e'), \mathbf{C_1}(e')\} \perp\!\!\!\perp M(\mathbf{c_1}, e)|\mathbf{C_0}$*.*

Theorem 1.1, in conjunction with the standard g-formula result $\delta_0 = \int_{\mathbf{c_0}} \mathbb{E}(Y|e', \mathbf{c_0}) dF(\mathbf{c_0})$ (Robins, 1986), which holds under the assumption encoded on the diagram that $Y(e') \perp\!\!\!\perp E|\mathbf{C_0}$, identifies the $\mathcal{P}_{EMY}$-specific effect, $\beta_0 - \delta_0$.

1.4 Path-specific inference

Thus far, we have only considered a nonparametric model $\mathcal{M}_{nonpar}$ for the observed data, making our identifying functional of the $\mathcal{P}_{EMY}$-specific effect valid under any possible correct model for the data. Unfortunately, we will seldom have the luxury to continue using $\mathcal{M}_{nonpar}$ through the estimation stage; because inference in $\mathcal{M}_{nonpar}$ is rarely practical in situations with numerous or continuous confounders $(\mathbf{C_0}, \mathbf{C_1})$ (Robins et al., 1997), we will often be forced to posit parametric models. Which models we are to fit depend on how we choose to estimate (1.1). We now consider four estimators and the corresponding models needed to compute them. Note that, while these estimators are in fact asymptotically equivalent under a nonparametric model, they will have different asymptotic properties under parametric and semiparametric models (Tchetgen Tchetgen and Shpitser, 2012).

1.4.1 Maximum Likelihood Estimation

We first discuss the maximum likelihood estimator (MLE) for β_0. By considering the identifying functional (1.1) as four nested expectations, it is clear that we can fit three appropriate regression models with parameters $\boldsymbol{\gamma_1}$, $\boldsymbol{\gamma_2}$, and $\boldsymbol{\gamma_3}$ using maximum likelihood, and plug the predicted means under these models into the functional; the outermost mean can then be estimated empirically. If the conditional mean of Y is taken to be linear in M and $\mathbf{C_1}$, and the conditional mean of M is taken as linear in $\mathbf{C_1}$, then mean models can be fit for Y, M, and $\mathbf{C_1}$. Thus, the MLE is

$$\hat{\beta}_{mle} \equiv \mathbb{P}_n \left\{ \hat{\mathbb{E}}(\hat{\mathbb{E}}(\hat{\mathbb{E}}(Y|M, \mathbf{C_1}, e', \mathbf{C_0}; \hat{\boldsymbol{\gamma}}_\mathbf{1})|\mathbf{C_1}, e, \mathbf{C_0}; \hat{\boldsymbol{\gamma}}_\mathbf{2})|e', \mathbf{C_0}; \hat{\boldsymbol{\gamma}}_\mathbf{3}) \right\},$$

where $\mathbb{P}_n$ denotes the empirical mean.

Define $\boldsymbol{\gamma} \equiv (\boldsymbol{\gamma_1}, \boldsymbol{\gamma_2}, \boldsymbol{\gamma_3})$, $g(\boldsymbol{\gamma}) \equiv \hat{\mathbb{E}}(\hat{\mathbb{E}}(\hat{\mathbb{E}}(Y|M, \mathbf{C_1}, e', \mathbf{C_0}; \boldsymbol{\gamma_1})|\mathbf{C_1}, e, \mathbf{C_0}; \boldsymbol{\gamma_2})|e', \mathbf{C_0}; \boldsymbol{\gamma_3})$, $\mathbf{D}_\gamma \equiv \mathbb{E}[\nabla_\gamma g(\boldsymbol{\gamma})]$, and $\mathbf{U}(\boldsymbol{\gamma})$ and $\boldsymbol{\mathcal{I}}(\boldsymbol{\gamma})$ to be the vector of score equations and block-diagonal matrix of expected informations, respectively, for $\boldsymbol{\gamma}$. Let $\boldsymbol{\gamma_0}$ be the true value of $\boldsymbol{\gamma}$. Then $\hat{\beta}_{mle}$ is asymptotically normal with asymptotic variance equal to $\mathbb{E}[(g(\boldsymbol{\gamma_0}) + \mathbf{D}_{\gamma_0}^T \boldsymbol{\mathcal{I}}(\boldsymbol{\gamma_0})\mathbf{U}(\boldsymbol{\gamma_0}) - \beta_0)^2]$, which can be estimated empirically, substituting $\hat{\gamma}$ and $\hat{\beta}_{mle}$ for $\boldsymbol{\gamma_0}$ and β_0. The MLE is asymptotically efficient when the three regression models are correctly specified, hence this is the minimum variance achievable by regular, asymptotically linear estimators under the choice of model $\mathcal{M}_{par}$ of $\mathbf{O}$. $\hat{\beta}_{mle}$ will be consistent only under correct specification of the three models.

1.4.2 Multiply-Robust Estimation

The multiply-robust (MR) estimator, $\hat{\beta}_{mr}$, comes from an estimating equation involving the efficient influence function of β_0 in the model $\mathcal{M}_{nonpar}$ placing no restriction on the observed data likelihood apart from the positivity assumptions given above. A derivation of this influence function is given in the supplementary materials. In order to express the estimator more succinctly, we introduce additional notation: $B(m, \mathbf{c_1}, e', \mathbf{c_0}) \equiv \mathbb{E}(Y|m, \mathbf{c_1}, e', \mathbf{c_0})$, $B'(\mathbf{c_1}, e', e, \mathbf{c_0}) \equiv \mathbb{E}\{\mathbb{E}(Y| M, \mathbf{c_1}, e', \mathbf{c_0})|\mathbf{c_1}, e, \mathbf{c_0}\}$, $B''(e', e, \mathbf{c_0}) \equiv \mathbb{E}[\mathbb{E}\{\mathbb{E}(Y|M, \mathbf{C_1}, e', \mathbf{c_0})|\mathbf{C_1}, e, \mathbf{c_0}\}|e', \mathbf{c_0}]$, $M^{ratio} \equiv f($

$M|\mathbf{C_1}, e, \mathbf{C_0})/f(M|\mathbf{C_1}, e', \mathbf{C_0})$, and $C_1^{ratio} \equiv f(\mathbf{C_1}|e, \mathbf{C_0})/f(\mathbf{C_1}|e', \mathbf{C_0})$. The estimator is then

$$\begin{aligned}\hat{\beta}_{mr} =& \mathbb{P}_n \Bigg\{ \frac{1_{e'}(E)}{\hat{f}(e'|\mathbf{C_0})} \hat{M}^{ratio} \Big\{ Y - \hat{B}(M, \mathbf{C_1}, e', \mathbf{C_0}) \Big\} \\ &+ \frac{1_e(E)}{\hat{f}(e|\mathbf{C_0})} (\hat{C}_1^{ratio})^{-1} \Big\{ \hat{B}(M, \mathbf{C_1}, e', \mathbf{C_0}) - \hat{B}'(\mathbf{C_1}, e', e, \mathbf{C_0}) \Big\} \\ &+ \frac{1_{e'}(E)}{\hat{f}(e'|\mathbf{C_0})} \Big\{ \hat{B}'(\mathbf{C_1}, e', e, \mathbf{C_0}) - \hat{B}''(e', e, \mathbf{C_0}) \Big\} + \hat{B}''(e', e, \mathbf{C_0}) \Bigg\},\end{aligned}$$

where $1_{e^*}(\cdot)$ is the indicator function, $\hat{B}'(\mathbf{C_1}, e', e, \mathbf{C_0}) = \hat{\mathbb{E}}\{\hat{B}(M, \mathbf{C_1}, e', \mathbf{C_0})|\mathbf{C_1}, e, \mathbf{C_0}\}$, and $\hat{B}''(e', e, \mathbf{C_0}) = \hat{\mathbb{E}}[\hat{\mathbb{E}}\{\hat{B}(M, \mathbf{C_1}, e', \mathbf{C_0})|\mathbf{C_1}, e, \mathbf{C_0}\}|e', \mathbf{C_0}]$.

Note that the estimator is only a function of estimates of $f_{M|\mathbf{C_1},E,\mathbf{C_0}}$ and $f_{\mathbf{C_1}|E,\mathbf{C_0}}$ through the ratios M^{ratio} and C_1^{ratio} and mean functions $B'(\mathbf{C_1}, e', e, \mathbf{C_0})$ and $B''(e', e, \mathbf{C_0})$. When the mean of Y is linear in M, then $B'(\mathbf{C_1}, e', e, \mathbf{C_0})$ only depends on the distribution of M through its conditional mean, $\mathbb{E}(M|\mathbf{C_1}, e, \mathbf{C_0})$. Similarly, if in addition the means of Y and M are both linear in $\mathbf{C_1}$, then $B''(e', e, \mathbf{C_0})$ only depends on the distribution of $\mathbf{C_1}$ through its conditional mean, $\mathbb{E}(\mathbf{C_1}|e', \mathbf{C_0})$. We denote $\boldsymbol{\theta_M} \equiv \{B'(\mathbf{C_1}, e', e, \mathbf{C_0}), M^{ratio}\}$ and $\boldsymbol{\theta}_{\mathbf{C_1}} \equiv \{B''(e', e, \mathbf{C_0}), C_1^{ratio}\}$.

B, $\boldsymbol{\theta_M}$, $\boldsymbol{\theta}_{\mathbf{C_1}}$, and $f_{E|\mathbf{C_0}}$ are estimated using low dimensional parametric working models, B^W, $\boldsymbol{\theta_M^W} = \{\mathbb{E}^W[B^W(M, \mathbf{C_1}, e', \mathbf{C_0})|\mathbf{C_1}, e, \mathbf{C_0}], M^{ratio;W}\}$, $\boldsymbol{\theta_{C_1}^W} = \{\mathbb{E}^W[B'^W(\mathbf{C_1}, e, \mathbf{c_0})|e', \mathbf{C_0}]$, $C_1^{ratio:W}\}$, and $f^W_{E|\mathbf{C_0}}$, via standard maximum likelihood. Note that we are able to avoid estimating the densities for $\mathbf{C_1}$ and M by instead estimating their mean functions and density ratios directly. Mean functions can be estimated with standard regression techniques, and density ratios can be estimated using propensity score models since by Bayes' theorem,

$$\frac{f(\mathbf{C_1}|e, \mathbf{C_0})}{f(\mathbf{C_1}|e', \mathbf{C_0})} = \frac{f(e|\mathbf{C_1}, \mathbf{C_0})}{f(e'|\mathbf{C_1}, \mathbf{C_0})} \times \frac{f(e'|\mathbf{C_0})}{f(e|\mathbf{C_0})}$$

and

$$\frac{f(M|e, \mathbf{C_1}, \mathbf{C_0})}{f(M|e', \mathbf{C_1}, \mathbf{C_0})} = \frac{f(e|M, \mathbf{C_1}, \mathbf{C_0})}{f(e'|M, \mathbf{C_1}, \mathbf{C_0})} \times \frac{f(e'|\mathbf{C_1}, \mathbf{C_0})}{f(e|\mathbf{C_1}, \mathbf{C_0})}.$$

An attractive property of the multiply-robust estimator is its robustness to multiple types of potential model misspecification. Let $\hat{B}$, $\hat{\boldsymbol{\theta}}_{\mathbf{M}}$, $\hat{\boldsymbol{\theta}}_{\mathbf{C_1}}$, and $\hat{f}_{E|\mathbf{C_0}}$ denote estimators of B^W, $\boldsymbol{\theta_M^W}$, $\boldsymbol{\theta_{C_1}^W}$, and $f^W_{E|\mathbf{C_0}}$ consistent under correct specification. The mean functions in $\boldsymbol{\theta_M}$ and $\boldsymbol{\theta}_{\mathbf{C_1}}$ require correct specification of the functions of M and $\mathbf{C_1}$ based on the working models for Y and $\{M, Y\}$, respectively, so that $\boldsymbol{\theta_M^W}$ and $\boldsymbol{\theta_{C_1}^W}$ can be correctly specified regardless of whether B^W is, and

$\boldsymbol{\theta}^{\boldsymbol{W}}_{\boldsymbol{C_1}}$ can be correctly specified regardless of whether $\boldsymbol{\theta}^{\boldsymbol{W}}_{\boldsymbol{M}}$ is. The multiply-robust estimator is consistent and asymptotically normal (under standard regularity conditions) provided that one of the following holds: (a) $\{\boldsymbol{\theta_M}, f_{E|\mathbf{C_0}}\} \in \{\boldsymbol{\theta^W_M}, f^W_{E|\mathbf{C_0}}\}$, (b) $\{B, \boldsymbol{\theta}_{\mathbf{C_1}}, f_{E|\mathbf{C_0}}\} \in \{B^W, \boldsymbol{\theta^W_{C_1}}, f^W_{E|\mathbf{C_0}}\}$, (c) $\{B, \boldsymbol{\theta}_{\mathbf{C_1}}, \boldsymbol{\theta_M}\} \in \{B^W, \boldsymbol{\theta^W_{C_1}}, \boldsymbol{\theta^W_M}\}$. That is, $\hat{\beta}_{mr}$ offers three distinct opportunities to obtain valid inference about the path-specific effect. By contrast, $\hat{\beta}_{mle}$ will be consistent only if a slightly weaker form of (c) holds, where $M^{ratio;W}$ and $C_1^{ratio;W}$ need not be correctly specified.

For inference on $\hat{\beta}_{mr}$, we recommend the nonparametric bootstrap (Efron, 1979) or similar alternative resampling methods such as the randomly weighted bootstrap (Rao and Zhao, 1992; Van Der Vaart and Wellner, 1996) for nonparametric variance estimation. Due to its reliance on inverse-propensity-score weights, this estimator may suffer from instability in settings where the set of positivity assumptions is nearly violated (Kang and Schafer, 2007). A useful stabilization technique is to simply replace any propensity score $\hat{f}_{E|\mathbf{X}}$ with $\hat{f}^{\dagger}_{E|\mathbf{X}}$, where $\mathbf{X}$ is some vector of covariates and $\text{logit}\hat{f}_{E|\mathbf{X}}(e|\mathbf{X}) = \text{logit}\hat{f}_{E|\mathbf{X}}(e|\mathbf{X}) - \log(1 - \mathbb{P}_n(1_e(E))) + \log(\mathbb{P}_n[1_e(E)\hat{f}_{E|\mathbf{X}}(e'|\mathbf{X})/\hat{f}_{E|\mathbf{X}}(e|\mathbf{X})])$, which ensures the weights are bounded as discussed in Tchetgen Tchetgen and Shpitser (2012). An additional stabilization technique is given in the supplementary materials.

1.4.3 Other Estimators

We consider two additional estimators, both based on alternative representations of (1.1) as shown in the supplementary materials:

$$\hat{\beta}_a \equiv \mathbb{P}_n \left\{ \frac{1_{e'}(E)}{\hat{f}(e'|\mathbf{C_0})} \hat{M}^{ratio} Y \right\}$$

$$\hat{\beta}_b \equiv \mathbb{P}_n \left\{ \frac{1_e(E)}{\hat{f}(e|\mathbf{C_0})} (\hat{C}_1^{ratio})^{-1} \hat{\mathbb{E}}(Y|M, \mathbf{C_1}, e', \mathbf{C_0}) \right\},$$

which again involve plugging in estimated regression models and density curves $\hat{f}(e|M, \mathbf{C_1}, \mathbf{C_0})$, $\hat{f}(e|\mathbf{C_1}, \mathbf{C_0})$, and $\hat{f}(e|\mathbf{C_0})$. Note that $\hat{\beta}_a$ and $\hat{\beta}_b$ depend only on a subset of the models in the multiple-robustness conditions (a) and (b), respectively. It follows that $\hat{\beta}_a$ will generally be consistent only if a slightly weaker form of (a) holds, where B'^W need not be correctly specified. Similarly, $\hat{\beta}_b$ will be consistent only if a slightly weaker form of (b) holds, where B''^W need not be correctly specified. In settings with practical violations of positivity, stability of both estimators can be improved using the stabilization technique given in Section 1.4.2.

1.5 Simulation study

We report results for a simulation study in which we generated 1000 data sets of size 1000 from the following models:

$$
\begin{aligned}
C_0 &\sim \mathcal{U}(0,2) \\
E|C_0 &\sim Bernoulli\left(1-(1+\exp(0.9+0.3C_0))^{-1}\right) \\
\mathbf{C_1} &= \begin{pmatrix} 0.8 \\ 0.6 \\ -0.3 \end{pmatrix} + \begin{pmatrix} 1 \\ 0.1 \\ 0.2 \end{pmatrix} C_0 + \begin{pmatrix} 0.5 \\ -0.4 \\ 0.5 \end{pmatrix} E + \begin{pmatrix} -0.1 \\ 0.8 \\ -0.2 \end{pmatrix} C_0 E + \mathcal{N}(0,I) \\
M &= -0.5 - 0.2C_0 + 0.3E + [-0.2, 0.1, 0.5]\mathbf{C_1} + [0.4, 0, 0]E\mathbf{C_1} + N(0,1) \\
Y &= 0.2 + 0.2C_0 + 0.6E + [1, 0.7, 0.3]\mathbf{C_1} - 0.9M - 0.8EM + N(0,1).
\end{aligned}
$$

In order to investigate the impact of model misspecification, we computed each of the four estimators given above, $\hat{\beta}_{mr}$, $\hat{\beta}_{mle}$ $\hat{\beta}_a$, and $\hat{\beta}_b$, under the four parametric models, $\mathcal{M}_a$, $\mathcal{M}_b$, $\mathcal{M}_c$, and $\mathcal{M}_{int}$. Models $\mathcal{M}_a$, $\mathcal{M}_b$, and $\mathcal{M}_c$ were specified such that statements (a)-(c) in Section 1.4.2 corresponding to their respective subscripts held, but the models for the remaining estimands were incorrectly specified. For instance, under $\mathcal{M}_a$, models $\boldsymbol{\theta}^{\boldsymbol{W}}_{\boldsymbol{M}}$ and $f^W_{E|C_0}$ are correctly specified, while B^W and $\boldsymbol{\theta}^{\boldsymbol{W}}_{\boldsymbol{C_1}}$ are not. The intersection model uses correctly-specified working models. All models were fit by maximum likelihood. The stabilization technique described in Section 1.4.2 was used to adjust propensity scores. We used the following working models, subscripted C for correctly specified and I for incorrectly specified:

$f^W_{E|C_0}$:

Correct: logit $\Pr_C\{E=1|C_0\} = [1, C_0]\boldsymbol{\alpha_C}$

Incorrect: $\Phi^{-1}(\Pr_I\{E=1|C_0\}) = [1, C_0]\boldsymbol{\alpha_I}$

B^W:

Correct: $\mathbb{E}_C[Y|M, \mathbf{C_1}, E, C_0] = [1, C_0, E, \mathbf{C_1}, M, EM]\boldsymbol{\eta_C}$

Incorrect: $\mathbb{E}_I[Y|M, \mathbf{C_1}, E, C_0] = [1, C_0, E, \mathbf{C_1}, M]\boldsymbol{\eta_I}$

$\boldsymbol{\theta}^{\boldsymbol{W}}_{\boldsymbol{C_1}}$:

Correct: $C_1^{ratio;W} = \Pr_C(E=e|\mathbf{C_1}, C_0)/\Pr_C(E=e'|\mathbf{C_1}, C_0) \times \Pr_C(E=e'|C_0)\Pr_C(E=e|C_0)$, which depends on the correctly-specified $f^W_{E|C_0}$ model and the correctly-specified model logit $\Pr_C\{E=1|\mathbf{C_1}, C_0\} = [1, C_0, C_0^2, \mathbf{C_1}, C_0\mathbf{C_1}]\boldsymbol{\lambda_C}$;

$B''_C(e', e, C_0) = \mathbb{E}_C[\mathbb{E}_C\{\mathbb{E}_C(Y|M, \mathbf{C_1}, e', C_0)|\mathbf{C_1}, e, C_0\}|e', C_0]$, which depends on the correctly-specified B^W model and the correctly-specified models $\mathbb{E}_C[C_{1j}|E, C_0] = [1, C_0, E, C_0E]\boldsymbol{\delta_{j;C}}\ \forall j \in \{1, 2, 3\}$ and $\mathbb{E}_C[M|\mathbf{C_1}, E, C_0] = [1, C_0, E, \mathbf{C_1}, EC_{11}]\boldsymbol{\zeta_C}$.

Incorrect: $C_1^{ratio;W,I} = \text{Pr}_I(E = e|\mathbf{C_1}, C_0)/\text{Pr}_I(E = e'|\mathbf{C_1}, C_0) \times \text{Pr}_C(E = e'|C_0)/\text{Pr}_C(E = e|C_0)$, which depends on the correctly-specified $f^W_{E|C_0}$ model and the incorrectly-specified model logit $\text{Pr}_I\{E = 1|\mathbf{C_1}, C_0\} = [1, C_0, \mathbf{C_1}]\boldsymbol{\lambda_I}$;
$B''_I(e', e, C_0) = \mathbb{E}_I[\mathbb{E}_C\{\mathbb{E}_I(Y|M, \mathbf{C_1}, e', C_0)|\mathbf{C_1}, e, C_0\}|e', C_0]$, which depends on the incorrectly-specified B^W model, the correctly-specified working mean model for M used for $B''_C(e', e, C_0)$ above, and the incorrectly-specified model $\mathbb{E}_I[C_{1j}|E, C_0] = [1, C_0, E]\boldsymbol{\delta_{j,I}}$, since $\boldsymbol{\theta^W_{C_1}}$ is only misspecified in setting (a), under which B^W is also misspecified and $\boldsymbol{\theta^W_M}$ is correctly specified.
$\boldsymbol{\theta^W_M}$:

Correct: $M^{ratio;W,C} = \text{Pr}_C(E = e|M, \mathbf{C_1}, C_0)/\text{Pr}_C(E = e'|M, \mathbf{C_1}, C_0) \times \text{Pr}_C(E = e'|\mathbf{C_1}, C_0)/\text{Pr}_C(E = e|\mathbf{C_1}, C_0)$, which depends on the correctly-specified model logit $\text{Pr}_C\{E = 1|M, \mathbf{C_1}, C_0\} = [1, C_0, C_0^2, \mathbf{C_1}, C_0\mathbf{C_1}, C_{11}\mathbf{C_1}, M, C_{11}M]\boldsymbol{\gamma_C}$ and the correctly-specified logistic model used for $C_1^{ratio;W}$ above;
$B'_C(\mathbf{C_1}, e', e, C_0) = \mathbb{E}_C\{\mathbb{E}_C(Y|M, \mathbf{C_1}, e', C_0)|\mathbf{C_1}, e, C_0\}$ depends on the correctly-specified B^W model and the correctly-specified mean model for M used for $B''_C(e', e, C_0)$ above.

Incorrect: $M^{ratio;W,I} = \text{Pr}_I(E = e|M, \mathbf{C_1}, C_0)/\text{Pr}_I(E = e'|M, \mathbf{C_1}, C_0) \times \text{Pr}_C(E = e'|\mathbf{C_1}, C_0)/\text{Pr}_C(E = e|\mathbf{C_1}, C_0)$, which depends on the correctly-specified logistic model for $\text{Pr}_C\{E = 1|\mathbf{C_1}, C_0\}$ and the incorrectly-specified model logit $\text{Pr}_I\{E = 1|M, \mathbf{C_1}, C_0\} = [1, C_0, \mathbf{C_1}, M]\boldsymbol{\gamma_I}$;
$B'_I(\mathbf{C_1}, e', e, C_0) = \mathbb{E}_I\{\mathbb{E}_C(Y|M, \mathbf{C_1}, e', C_0)|\mathbf{C_1}, e, C_0\}$, which depends on the incorrectly-specified model $\mathbb{E}_I[M|\mathbf{C_1}, E, C_0] = [1, C_0, E, \mathbf{C_1}]\boldsymbol{\zeta_I}$ and the correctly-specified model $B^{W,C}$, since $\boldsymbol{\theta^W_M}$ is only misspecified in setting (c), under which B^W is correctly specified.

The results are summarized in the plot displayed in Figure 1.2 that shows the four point estimates under each model and their corresponding 95% confidence intervals. The point estimates are the Monte Carlo means of the 1000 samples and the confidence intervals are the values within $t_{999,0.975}$ times the corresponding Monte Carlo standard errors of the point estimates. The confidence intervals correspond to t tests of $H_0 : \hat{\beta} = \beta_0 \equiv 2.678$, hence the confidence intervals not containing β_0, represented by the horizontal dashed line, correspond to rejection of H_0.

All estimators are consistent under $\mathcal{M}_{int}$. Besides $\hat{\beta}_{mr}$, $\hat{\beta}_a$ is the only consistent estimator

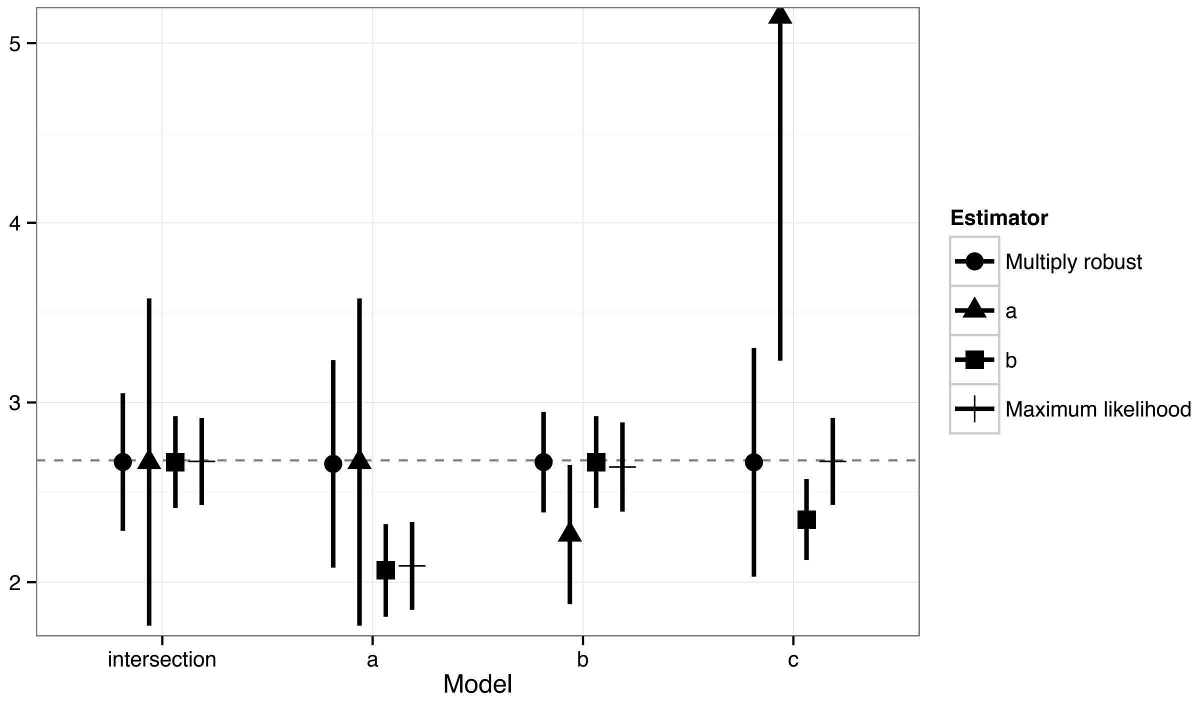


Figure 1.2: Simulation results for n=1000. Monte Carlo point estimates and confidence intervals of each of the four $\mathcal{P}_{EMY}$-specific estimators are given under $\mathcal{M}_{int}$, $\mathcal{M}_a$, $\mathcal{M}_b$, and $\mathcal{M}_c$. The horizontal dashed line is through the true parameter value, β_0.

under $\mathcal{M}_a$, $\hat{\beta}_b$ is the only consistent estimator under $\mathcal{M}_b$, and $\hat{\beta}_{mle}$ is the only consistent estimator under $\mathcal{M}_c$. $\hat{\beta}_{mr}$ is consistent under all models. Therefore, in moderate to large samples, we expect to reject H_0 at the nominal $\alpha = 0.05$ level for none of the estimators under $\mathcal{M}_{int}$, only for $\hat{\beta}_b$ and $\hat{\beta}_{mle}$ under $\mathcal{M}_a$, only for $\hat{\beta}_a$ and $\hat{\beta}_{mle}$ under $\mathcal{M}_b$, and only for $\hat{\beta}_a$ and $\hat{\beta}_b$ under $\mathcal{M}_c$.

The results illustrate quite well the multiple-robustness property of $\hat{\beta}_{mr}$. As predicted, while the other estimators failed to estimate β_0 without statistically-significant bias, the tests for $\hat{\beta}_{mr}$ failed to reject under every model. For the other estimators, the tests never rejected under $\mathcal{M}_{int}$ and their corresponding models where the misspecified components did not factor into estimation, as expected. That is, the test for $\hat{\beta}_a$ did not reject under $\mathcal{M}_a$, the test for $\hat{\beta}_b$ did not reject under $\mathcal{M}_b$, and the test for $\hat{\beta}_{mle}$ did not reject under $\mathcal{M}_c$. The tests did reject, however, under the other models, with the exception of $\hat{\beta}_c$ under $\mathcal{M}_b$. Thus, all estimators other than $\hat{\beta}_{mr}$ were significantly biased under at least one model.

We do see a tradeoff between efficiency and robustness; in all settings, $\hat{\beta}_{mle}$ and $\hat{\beta}_b$ perform best in terms of efficiency, with a slight advantage going to $\hat{\beta}_{mle}$. As such, a reasonable strategy may be to use these estimators in concert initially to diagnose model specification, and then select the most efficient estimator that appears to agree with the multiply-robust estimator, possibly $\hat{\beta}_{mr}$ itself.

Other sample sizes were also explored. At $n = 500$, asymptotic results began to come into focus, though a few of the tests expected to reject looked to be slightly underpowered due to the sample not being large enough. Still, the multiply-robust estimator outperformed the other estimators at this sample size in terms of robustness. At $n = 5000$, confidence intervals were tighter, as expected, but t test results were the same as those at $n = 1000$. The simulation study with $n = 1000$ is comparable to our data analysis in terms of sample size; every treatment comparison consisted of at least 1000 patients.

1.6 Harvard PEPFAR Nigeria analysis

We now present results of the Harvard PEPFAR data analysis. The data set consisted of 9968 complete observations, i.e., observations with no missing variables, which was 41.9% of the entire data set. We first consider the path-specific effect of treatment regimen assignment on virologic failure through adherence, expressed on the log-risk ratio scale. We used each of the four estimators for β_0 from Section 1.4, and in each case, δ_0 was estimated using only a subset of the models used to estimate β_0. In particular, the doubly-robust estimator (Bang and Robins, 2005) was used to estimate δ_0 when contrasted with $\hat{\beta}_b$ and the multiply-robust estimator, $\hat{\beta}_{mr}$; the inverse-probability-weighted estimator (IPW) (Horvitz and Thompson, 1952) was used to estimate δ_0 when contrasted with $\hat{\beta}_a$; and the MLE was used to estimate δ_0 when contrasted with the MLE for the β_0, $\hat{\beta}_{mle}$. Accordingly, let $\hat{\mathcal{P}}_{EMY;mle}$ denote the effect estimate using $\hat{\beta}_{mle}$, $\hat{\mathcal{P}}_{EMY;a}$ denote the effect estimate using $\hat{\beta}_a$, $\hat{\mathcal{P}}_{EMY;b}$ denote the effect estimate using $\hat{\beta}_b$, and $\hat{\mathcal{P}}_{EMY;mr}$ denote the effect estimate using $\hat{\beta}_{mr}$. We computed all four estimates and corresponding bootstrap confidence intervals for each pairwise comparison of treatments. A randomly weighted bootstrap (Rao and Zhao, 1992; Van Der Vaart and Wellner, 1996) with weights sampled from $\mathrm{Exp}(1)$ was used to account for instability of resamples due to a small number of cases in some strata of E. Results are summarized in Figure C.1 of the supplementary materials.

$\hat{\mathcal{P}}_{EMY;a}$ agreed with $\hat{\mathcal{P}}_{EMY;mr}$ across all comparisons, suggesting that it did not suffer much from model misspecification, or at least any worse than did $\hat{\mathcal{P}}_{EMY;mr}$. It also proved to be the more efficient estimator in this setting, with confidence intervals that were narrower than those of $\hat{\mathcal{P}}_{EMY;mr}$, and comparable to those of $\hat{\mathcal{P}}_{EMY;mle}$, which did not appear to be as robust. Thus, we chose to perform inference using $\hat{\mathcal{P}}_{EMY;a}$ for this portion of the analysis.

Recall that the treatment regimens were coded in descending order of magnitude of their total effects on risk of virologic failure, i.e., they were coded in ascending order of counterfactual risk of virologic failure had everyone been assigned to that treatment, since a lower counterfactual risk of failure corresponds to a higher magnitude of total effect. Because in practice we are more interested in learning how less-effective treatments can be improved, we only consider the higher-coded treatment in a pair as the baseline, e'. Using this ordering, the path-specific effect gives the improvement over the total effect of the less-effective treatment when intervening to make patients adhere as if they were on the more-effective treatment, but had the toxicity and direct effectiveness of the less-effective treatment.

We are primarily interested in the proportion of the total effect attributable to the mediated effect, i.e., the percent mediated by $\mathcal{P}_{EMY}$. If this proportion is close to or exceeds one, we can conclude that the drugs themselves likely have the same effectiveness on virologic failure, and that it is their differential effect on adherence not due to toxicity that is driving the difference in total effects. If, on the other hand, this proportion is small or negative, we can only say that the difference in total effects is not driven by a difference in effects through $\mathcal{P}_{EMY}$. It may be the case that the efficacies of the drugs themselves do, in fact, differ, or that the difference in total effects is driven by the differential effect on adherence due to toxicity, but we cannot confirm either. Table 1.2 shows $\hat{\mathcal{P}}_{EMY;a}$ divided by the total effect estimates, which are also on the log-risk ratio scale and are estimated by IPW. Superscripts indicate the comparisons with significant and marginally-significant path-specific effects. Due to the treatment coding, the denominators of the Table 1.2 values are always negative. Thus, a negative path-specific effect will be in the same direction as the total effect, and hence will explain a positive proportion of it.

Note that all significant and marginally-significant proportions of total effects due to the effects through $\mathcal{P}_{EMY}$ were negative apart from the one comparing treatment 1 (TDF+3TC/FTC+EFV) to baseline treatment 2 (d4T+3TC+NVP). This occurs when the $\mathcal{P}_{EMY}$-specific effect estimate

Table 1.2: Proportion of total effect on virologic failure due to $\mathcal{P}_{EMY}$-specific effect

	Baseline treatment			
Comparison trt	2	3	4	5
1	0.41†	0.21	-0.059	-0.068*
2	-	0.13	-0.49*	-0.20*
3	-	-	-0.57†	-0.13*
4	-	-	-	-0.027†

NOTE: *Significant path-specific effect ($\alpha = 0.05$). †Marginally-significant path-specific effect ($\alpha = 0.1$).

is in the opposite direction of the total effect estimate, suggesting that directionally-opposite effects through other pathways overwhelm our estimated effect, and that the total effect would have been even greater if not for the $\mathcal{P}_{EMY}$-specific effect. For example, had the $\mathcal{P}_{EMY}$-specific effect been null in the case comparing treatment 3 (AZT+3TC+EFV) with treatment 5 (TDF+3TC/FTC+NVP), we estimate that the total effect would have been 13% larger. The effect of treatment 3 is stronger than 5 not because of its effect through $\mathcal{P}_{EMY}$, but in spite of it. All differences between treatment 5 and another treatment, and all differences between treatment 4 (AZT+3TC+NVP) and another treatment besides 1 were observed to exhibit this phenomenon as well.

Now consider the exception noted above: the comparison of treatment 1 (TDF+3TC/FTC+EFV) to baseline treatment 2 (d4T+3TC+NVP). We saw a marginally-negative effect, which would have the following interpretation: the effect of treatment 2 on the risk of virologic failure would be improved by patients adhering as if they were assigned to treatment 1, but still had the same toxicity that they did on treatment 2. Unfortunately, treatment 2 is known to have toxicities that were not measured in this data set that are likely to also be affected by underlying biological causes of virologic failure. This interpretation cannot even be considered to be valid for the effect through these unmeasured toxicities, since they induce unmeasured confounding that once again renders this effect unidentifiable. If there were no unmeasured toxicities, we would interpret this effect as accounting for an estimated 41% of the differences in total effects between treatments 1 and 2.

In conclusion, of the significant $\mathcal{P}_{EMY}$-specific effects we observed, all apart from those involving unmeasured toxicities were countervailing to the total effect. This means that for these treatment pairs, the differences in their total effects on virologic failure would have been even greater if not for the effect along $\mathcal{P}_{EMY}$. Thus, the effect through $\mathcal{P}_{EMY}$ does not explain the differential effects on virologic failure, and in some cases actually works against them. As mentioned above, the differential effects may instead be due to the drugs themselves differing in efficacy, or they may be driven by the differential effects on adherence due to toxicity, but such hypotheses require further investigation beyond the scope of our analysis.

We now consider the path-specific effect of treatment regimen assignment on log CD4 count, expressed on the mean difference scale. We again analyzed the four estimators given in Section 1.4. This time $\hat{\mathcal{P}}_{EMY;a}$ and $\hat{\mathcal{P}}_{EMY;b}$ were drastically less efficient than $\hat{\mathcal{P}}_{EMY;mle}$ and $\hat{\mathcal{P}}_{EMY;mr}$. One possible explanation for this is that the density of log CD4 count was less concentrated around zero, making $\hat{\mathcal{P}}_{EMY;a}$ and $\hat{\mathcal{P}}_{EMY;b}$ more sensitive to small weights. $\hat{\mathcal{P}}_{EMY;mle}$ disagreed with $\hat{\mathcal{P}}_{EMY;mr}$ on several occassions, so $\hat{\mathcal{P}}_{EMY;mr}$ was the best choice in terms of achieving both robustness and efficiency. It is worth noting that the linear outcome model for CD4 count did not seem to suffer too much from misspecification, while the logistic outcome model for virologic failure did. Results are summarized in Figure C.2 of the supplementary materials.

Table 1.3 shows $\hat{\mathcal{P}}_{EMY;mr}$ divided by the total effect estimates, which are also on the log-risk ratio scale and are estimated using doubly-robust estimators. As before, we are interested in learning how the less-effective treatment can be improved, but now less-effective is in terms of

Table 1.3: Proportion of total effect on CD4 count due to $\mathcal{P}_{EMY}$-specific effect

	Baseline treatment			
Comparison trt	3	1	5	2
4	0.48*	0.095	-0.045†	0.036
3	-	-0.47	-0.13	0.030
1	-	-	-0.080	0.062
5	-	-	-	0.099

NOTE: *Significant path-specific effect ($\alpha = 0.05$). †Marginally-significant path-specific effect ($\alpha = 0.1$).

CD4 count. Since the order of the effectiveness of the treatments for CD4 count is not the same as the order for virologic failure, the treatments which should be considered the comparison versus baseline level in a pair no longer correspond to the treatment coding. The order of the treatments in the margins of Table 1.3 is rearranged to reflect this different ordering of effectiveness. The denominator for each of the values in the table is positive, since a higher counterfactual CD4 count corresponds to a higher magnitude of total effect. Therefore, positive proportions correspond to positive path-specific effects, and negative proportions correspond to countervailing path-specific effects.

The path-specific effect was found to be significant for only one of the pairwise comparisons: treatment 4 (AZT+3TC+NVP) vs. treatment 3 (AZT+3TC+EFV). This effect is estimated to be in the positive direction, therefore we conclude that the effect of treatment 3 on CD4 count would be improved by patients adhering as if they were assigned to treatment 4 but without necessarily altering toxicity experienced under treatment 3 that they did on treatment 3. The effect through this pathway accounted for almost half of the total effect at an estimated 48%. Thus, if one were interested in improving the effect of AZT+3TC+EFV on CD4 count, it would be worthwhile to examine what mechanisms other than toxicity may be implicated in differential adherence rates between these two regimens. The $\mathcal{P}_{EMY}$-specific effect comparing treatments 4 and 5 (TDF+3TC/FTC+NVP) was found to be marginally-significantly less than zero. Thus, the difference in total effects of these two treatments is not attributable to their differential effect on adherence not due to toxicity, as the effect through this pathway was in fact in the opposite direction. Rather, this difference was due to differential effects through other pathways as previously described.

1.7 Discussion

In the PEPFAR case study, we observed an interesting trend of countervailing effects along $\mathcal{P}_{EMY}$ to the total effects on virologic failure for most treatment comparisons, meaning that the differences in the total effects of treatment assignment would have been even greater if not for the effects along $\mathcal{P}_{EMY}$. While this does not help explain why the treatment assignment effects are different (or at least different in the direction that we observe), it does suggest a method for improving the regimens that we observed to have greater effects on virologic failure. For a treatment comparison with a significant $\mathcal{P}_{EMY}$-specific effect, if we could identify what is different about the more effective drug regimen that is causing people to not adhere to it as well, then we could potentially eliminate this mechanism in order to reduce the countervailing $\mathcal{P}_{EMY}$-specific effect and consequently improve its total effect on virologic failure.

A countervailing $\mathcal{P}_{EMY}$-specific effect on CD4 count was also observed between AZT+3TC+NVP and TDF+3TC/FTC+NVP, which has the same interpretation as the countervailing effects on virologic failure. On the other hand, almost half of the difference in the effects of AZT+3TC+EFV and AZT+3TC+NVP on CD4 count was found to be attributable to the effect through adherence, but not toxicity. This suggests that the effect of AZT+3TC+EFV on CD4 count could be improved up to that of AZT+3TC+NVP if one could identify and eliminate the mechanisms driving the difference in these treatments' effects on adherence. In the other treatment comparisons, none of the differences in total effects on CD4 count were found to be attributable to an effect through $\mathcal{P}_{EMY}$. Overall, we have achieved an enhanced understanding of the role of adherence in the effects of the five ART regimens considered on both virologic failure and CD4 count.

The most significant methodologic contribution of this paper is the extension of mediation analysis methods to settings in which the NDE and NIE may not be identified, viz. settings with unmeasured confounding and exposure-induced confounding of the mediator. We present conditions under which the $\mathcal{P}_{EMY}$-specific effect is nonparametrically identified as well as four estimators, including an efficient estimator that is multiply robust to model misspecification for settings where nonparametric estimation is not feasible.

Often effects of adherence are evaluated regarding the treatment assignment as an instrumental variable, relying on an assumption of no direct effect of assignment with respect to adherence. Furthermore, instrumental variable methods rely on an assumption of monotonicity in the effect of assignment on adherence. However, neither of these assumptions are reasonable in our setting where we are forced to compare treatments head-to-head rather than to a control exposure level.

This paper suffers from a few limitations. One is that our identifiability assumptions, though weaker than those of the Markovian model, are still untestable as stated. When possible, we can embed our mediation problem in a larger model represented by a larger graph where treatments can be split into a component corresponding to the EMY pathway and a component corresponding to all other pathways. This can provide a testable reformulation of identifying assumptions, as was done in Robins and Richardson (2010) in simpler mediation contexts. Another limitation is that this method is not yet equipped to handle missing data. As such, only a complete-case analysis was conducted for the HIV data, allowing for the possibility of bias due to informative missingness. Additionally, for both virologic failure and CD4 count outcomes, it is possible that we are underestimating the effect of substantive interest if adherence over the first six months plays a large mediating role since we are forced to control for early adherence and can only estimate the effect through adherence over the second six months. Finally, not a limitation, but rather a caveat, is that the $\mathcal{P}_{EMY}$-specific effect is not a substitute for the NIE. The NIE is not fully captured by this effect and, in fact, even if the effects along both $\mathcal{P}_{EMY}$ and $E \rightarrow \mathbf{C_1} \rightarrow M \rightarrow Y$ are in the same direction, the NIE does not necessarily have to be. Strong assumptions are needed to draw this conclusion. As such, while often practically meaningful, the $\mathcal{P}_{EMY}$-specific effect must be interpreted with care and not blindly substituted for the NIE.

Future directions for this work would, of course, include adjusting the method to account for missing data, which could improve the analysis conducted in this paper. Another important extension would be to the full longitudinal case, with repeated exposures, mediators, and confounders. Shpitser (2013) gives the identifying functional for the analog to the $\mathcal{P}_{EMY}$-specific effect in this setting, but no estimation strategy exists as of yet. Finally, it is not uncommon for a mediator to be measured with error, which tends to induce bias as shown by VanderWeele et al. (2012). It would be valuable to adapt the methods of Tchetgen Tchetgen and Lin (2012) for handling this problem to our setting. Alternatively, parametric approaches have been suggested (Valeri et al., 2014) that could also potentially be adapted.

On Partial Identification of the Pure Direct Effect

Caleb H. Miles, Ilya Shpitser, Phyllis Kanki, Seema Meloni, and Eric J. Tchetgen Tchetgen

Department of Biostatistics

Harvard School of Public Health

2.1 Introduction

Causal mediation analysis seeks to determine the role that an intermediate variable plays in "transmitting" the effect from an exposure to an outcome. An "indirect effect" refers to the effect that goes through the intermediate variable in mediation analysis; a "direct effect" is a measure of the effect that does not. The study of causal mediation has in recent years enjoyed an explosion in popularity (Robins and Greenland, 1992; Robins, 1999, 2003; Pearl, 2001; Avin et al., 2005; Taylor et al., 2005; Petersen et al., 2006; Ten Have et al., 2007; Goetgeluk et al., 2008; van der Laan and Petersen, 2008; VanderWeele, 2009, 2011; VanderWeele and Vansteelandt, 2009, 2010; Imai et al., 2010a,b; Tchetgen Tchetgen, 2011; Tchetgen Tchetgen and Shpitser, 2014, 2012; Shpitser, 2013; Tchetgen Tchetgen, 2013), not only in terms of theoretical developments, but also in practice, most notably in the fields of epidemiology and social sciences. This strand of work is based on ideas originating from Robins and Greenland (1992) and Pearl (2001) grounded in the language of potential outcomes (Splawa-Neyman et al., 1990; Rubin, 1974, 1978) to give a nonparametric definition of effects involved in mediation analysis, allowing for settings where certain interactions and nonlinearities may be present.

Consider an intervention which sets the exposure of interest for all persons in the population to one of two possible values, a reference value or an active value. The total effect of such an intervention corresponds to the change of the counterfactual outcome mean if the exposure were set to the active value compared with if it were set to the reference value. Robins and Greenland (1992) formalized the concept of effect decomposition of the total effect into direct and indirect effects by defining pure direct and indirect effects. Pearl (2001) relabeled these effects as natural direct and indirect effects. The natural direct effect corresponds to the change in the counterfactual outcome mean under an intervention which changes a person's exposure status from the reference value to the active value, while maintaining the person's mediator to the value it would have had under the exposure reference value. In contrast, the natural indirect effect corresponds to the change in the average counterfactual outcome under an intervention that sets a person's exposure value to the active value, while changing the value of the mediator from the value it would have had under the reference exposure value, to its value under the active exposure value.

Identification of these effects has been somewhat controversial as it requires assumptions that may be overly restrictive for many applications in the health sciences. First, identification invokes a so-called cross-world-counterfactuals-independence assumption, which by virtue of involving counterfactuals under conflicting interventions on the exposure, can neither be enforced experimentally nor tested empirically (Robins and Richardson, 2010). Secondly, a necessary assumption for identification rules out the presence of exposure-induced confounding of the mediator's effect on the outcome, even if all confounders are observed. While this assumption is in principle testable provided no unmeasured confounding, more often than not, post-exposure covariates are altogether ignored in routine application, in which case mediation analyses may be invalid. These issues have recently been considered, and some work has been done on partial or point identification under a weaker assumption. Specifically, on the one hand Robins and Richardson (2010) and Tchetgen Tchetgen and VanderWeele (2012) provide conditions for point identification of the pure direct effect when a confounder is directly affected by the exposure. On the other hand, Robins and Richardson (2010) give bounds for the pure direct effect for binary mediator without making the cross-world-counterfactual-independence assumptions, but assuming no exposure-induced confounding of the mediator-outcome relation, and Tchetgen Tchetgen and Phiri (2014) extend these bounds to account for exposure-induced confounding. We build on this previous work to provide a number of new nonparametric bounds for the pure direct effects allowing for a polytomous mediator when either (i) exposure-induced confounding is present, or (ii) one does not assume that cross-world counterfactuals of the mediating and outcome variables are independent, or (iii) both hold simultaneously.

We apply these bounds to data from the Harvard PEPFAR program in Nigeria, where we evaluate the extent to which the effects of antiretroviral therapy on virological failure are mediated by a patient's adherence. Results are sensitive to the choice of assumptions made, consequently, we counsel investigators employing these effects to exercise caution in considering the formula they use for identification and to explicitly state the assumptions required for them to be valid. Where assumptions are empirically untestable, they should be argued for on the basis of scientific understanding, and ideally the alternative should be explored by employing partial identification bounds given both here and elsewhere. Sensitivity analyses for ranges of plausible associations between cross-world counterfactuals remain undeveloped, but would be highly beneficial for practical use in such situations, and are fertile ground for future work. Our hope is that the work presented here

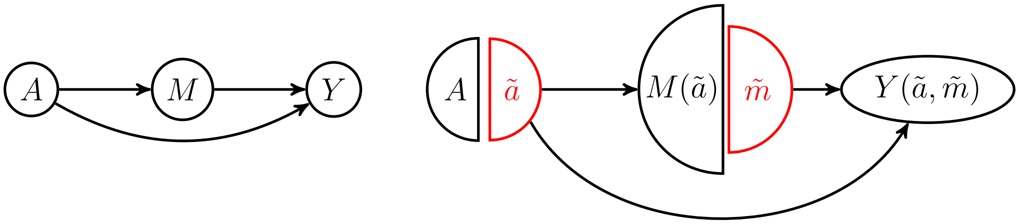


Figure 2.1: (a) The simple mediation directed acyclic graph in a setting with no confounding. (b) The single-world intervention graph in the setting of (a) that has been intervened on to set A to $\tilde{a}$ and M to $\tilde{m}$.

will inspire deeper consideration and transparency regarding underlying identifying assumptions in the practice of mediation analysis.

2.2 Preliminaries

By way of introduction, the directed acyclic graph (DAG) displayed in Figure 2.1(a) illustrates the simplest possible mediation setting, where A is defined to be the exposure taking either baseline value a^* or comparison value a, M is defined to be the (potential) mediator, and Y is defined to be the outcome. This DAG assumes randomization of the exposure, which for expositional simplicity we maintain throughout. The graph also encodes no unobserved confounding of the effect of M on Y. The effect along the path $A \rightarrow Y$ is considered direct with respect to M, and the effect along the path $A \rightarrow M \rightarrow Y$ is considered indirect with respect to M.

We now define counterfactual variables, letting $Y(a)$ denote a subject's outcome if treatment A were set, possibly contrary to fact, to a. In the context of mediation, there will also be potential outcomes for the intermediate variable. Counterfactuals $M(a)$ and $Y(m, a)$ are defined similarly. In order to link these with the observed data, we adopt the standard set of consistency assumptions that

if $A = a$, then $M(a) = M$ with probability one,

if $A = a$ and $M = m$, then $Y(m, a) = Y$ with probability one, and

if $A = a$, then $Y(a) = Y$ with probability one.

In terms of counterfactuals, the randomization assumption encoded by the DAG in Figure 2.1(a) is $\{Y(a, m), M(a)\} \perp\!\!\!\perp A$ for all a and m; the assumption of no unobserved confounding of M is $Y(a, m) \perp\!\!\!\perp M(a) \mid A$ for all a and m.

We may now define the pure/natural direct effect and natural indirect effect (Robins and Greenland, 1992; Pearl, 2001). These are expressed in terms of nested counterfactuals, i.e., counterfactual outcomes under an intervention which sets the exposure to a given value, and the mediator to the value it would have had under a possibly conflicting exposure value. They form the following decomposition of the average causal effect:

$$
\begin{aligned}
& E\left\{Y(a)\right\} - E\left\{Y(a^*)\right\} \\
&= \overbrace{E\left[Y\{a, M(a)\}\right] - E\left[Y\{a^*, M(a^*)\}\right]}^{\text{total effect}} \\
&= \overbrace{E\left[Y\{a, M(a)\}\right] - E\left[Y\{a, M(a^*)\}\right]}^{\text{natural indirect effect}} + \overbrace{E\left[Y\{a, M(a^*)\}\right] - E\left[Y\{a^*, M(a^*)\}\right)}^{\text{pure direct effect}}.
\end{aligned}
$$

The terms $E\{Y(a)\}$ and $E\{Y(a^*)\}$ are identified under randomization of A. The parameter $\gamma_0 \equiv E[Y\{a, M(a^*)\}]$ would be identified if one were to interpret the DAG in Figure 2.1(a) as a nonparametric structural equation model with independent errors (NPSEM-IE). Structural equations provide a nonparametric algebraic interpretation of this DAG corresponding to three equations, one for each variable in the graph. With each random variable on the graph is associated a distinct, arbitrary function, denoted g, and a distinct random disturbance, denoted ε, each with a subscript corresponding to its respective random variable. Each variable is generated by its corresponding function, which depends only on all variables that affect it directly (i.e., its parents on the graph), and its corresponding random disturbance, as follows:

$$\begin{aligned} A &= g_A(\varepsilon_A) \\ M &= g_M(A, \varepsilon_M) \\ Y &= g_Y(A, M, \varepsilon_Y). \end{aligned}$$

Under particular interventions, these structural equations naturally encode dependencies of counterfactuals. Consider, for example, two interventions, one setting $A = a^*$, and another setting $A = a$ and $M = m$. The structural equations then become

$$\begin{aligned} A &= a^* & A &= a \\ M(a^*) &= g_M(a^*, \varepsilon_M) & M(a) &= m \\ Y(a^*) &= g_Y(a^*, M(a^*), \varepsilon_Y) & Y(a, m) &= g_Y(a, m, \varepsilon_Y). \end{aligned}$$

This formulation places no restriction on the distribution of counterfactuals. The key assumption of the NPSEM-IE is that the random disturbances are mutually independent. This allows us to make independence statements regarding counterfactuals under various, possibly-conflicting interventions. In particular, this model implies that for all m, (i) $\{M(a), Y(a, m)\} \perp\!\!\!\perp A$, (ii) $Y(a, m) \perp\!\!\!\perp M \mid A = a$, and (iii) $Y(a, m) \perp\!\!\!\perp M(a^*) \mid A = a$, which in turn suffice for identification of γ_0 (Pearl, 2001). Independence statements such as (iii) are known as *cross-world counterfactual* statements due to their comparison of interventions that could never occur in the same world simultaneously. Independence (iii) can be seen to hold under the model by considering the NPSEM-IE under a specific intervention and noting that the only source of randomness in $Y(a, m) = g_Y(a, m, \varepsilon_Y)$ is ε_Y and the only source of randomness in $M(a^*) = g_M(a^*, \varepsilon_M)$ is ε_M. Thus, the cross-world-counterfactual-independence statement follows directly from independence of exogenous disturbances. However, such an independence is neither experimentally verifiable nor enforcable (Robins and Richardson, 2010).

This issue has been discussed extensively (Robins and Richardson, 2010; Richardson and Robins, 2013), and in large part motivated the development of the single-world intervention graphs (SWIGs) of Richardson and Robins (2013). These causal graphs manage to elucidate this issue by graphically representing the counterfactuals themselves, allowing independence statements of counterfactuals to be read directly from the graph. Consider the SWIG in Figure 2.1(b). By d-

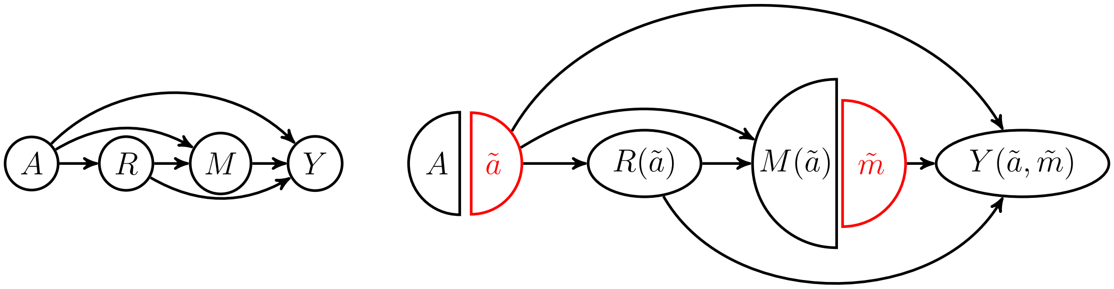


Figure 2.2: (a) A mediation directed acyclic graph in which R is an exposure-induced confounder. (b) The single-world intervention graph in the setting of (a) that has been intervened on to set A to $\tilde{a} \in \{a, a^*\}$ and M to $\tilde{m}$.

separation, it is clear that (i) $Y(a', m) \perp\!\!\!\perp M(a')$ for all a' and m, however under a model corresponding to the SWIG generated by the graph in Figure 2.1(a), no such statement can be made about $Y(a, m)$ and $M(a^*)$ when $a \neq a^*$. While γ_0 is not point identified under this model, Robins and Richardson (2010) provide the following bounds for its partial identification in the setting where M is binary and independence assumptions $M(a') \perp\!\!\!\perp A$ and $Y(a', m) \perp\!\!\!\perp \{M(a'), A\}$ hold for all a' and m:

$$\begin{aligned}
&\max\{0, \mathrm{pr}(M = 0 \mid A = a^*) + E(Y \mid M = 0, A = a) - 1\} \\
&\quad + \max\{0, \mathrm{pr}(M = 1 \mid A = a^*) + E(Y \mid M = 1, A = a) - 1\} \\
&\qquad\qquad \leq \gamma_0 \leq \\
&\min\{\mathrm{pr}(M = 0 \mid A = a^*), E(Y \mid M = 0, A = a)\} \\
&\quad + \min\{\mathrm{pr}(M = 1 \mid A = a^*), E(Y \mid M = 1, A = a)\}.
\end{aligned}$$

In Section 2.2, we extend this result to the setting of a polytomous M.

As previously mentioned, another often-overlooked condition required for identification of γ_0 is that there is no confounder of the mediator's effect on the outcome that is affected by the exposure. Such a confounder is present in the setting illustrated in the DAG in Figure 2.2(a). Generally, even under an NPSEM-IE , γ_0 will not be identified in this setting. This is readily seen by considering the following representation under this model given by Robins and Richardson

(2010):

$$\gamma_0 = \sum_{r,r^*} E\left\{E(Y \mid M, R=r, A=a) \mid R=r^*, A=a^*\right\} \mathrm{pr}\left\{R(a)=r, R(a^*)=r^*\right\}. \tag{2.1}$$

Clearly the joint probability term can never be identified from observed data, since we will never be able to observe $R(a)$ and $R(a^*)$ for the same individual.

A few conditions for identification have been proposed for binary R and M. Robins and Richardson (2010) give two. The first is that $R(a) \perp\!\!\!\perp R(a^*)$, in which case the troublesome term in (2.1) will factor, giving

$$\gamma_0 = \sum_{r^*,r} E\left\{E(Y \mid M, R=r, A=a) \mid R=r^*, A=a^*\right\} \mathrm{pr}(R=r^* \mid A=a^*) \\ \times \mathrm{pr}(R=r \mid A=a).$$

It seems unlikely, however, that in a scenario in which A affects R, the counterfacual R under $A = a$ would not be predictive of the counterfactual R under $A = a^*$. The other condition is that the counterfactual outcome under one exposure value is a deterministic function of the counterfactual for the other treatment, i.e., $R(a) = g\{R(a^*)\}$. In this case,

$$\gamma_0 = \sum_{r^*,r} E\left\{E(Y \mid M, R=r, A=a) \mid R=r^*, A=a^*\right\} \mathrm{pr}(R=r^* \mid A=a^*) I\{r=g(r^*)\}.$$

The above assumption is implied by rank preservation (Robins and Richardson, 2010), which is unlikely to hold in most social and health sciences as it rules out individual-level effect heterogeneity (Tchetgen Tchetgen and VanderWeele, 2012). As none of these conditions are experimentally verifiable, the authors themselves “do not advocate blithely adopting such assumptions in order to preserve identification of the PDE in [this setting]" (Robins and Richardson, 2010).

Tchetgen Tchetgen and VanderWeele (2012) give two testable conditions for identification of γ_0 when R is present. The first is of A–R monotonicity, i.e., for Bernoulli R, $R(a) \geq R(a^*)$. If R is a vector of Bernoulli random variables whose structural equations have independent errors, and if monotonicity holds for each element,

$$\gamma_0 = \sum_{r,r^*} E\left\{E(Y \mid M, R=r, A=a) \mid R=r^*, A=a^*\right\} \prod_{j=1}^{k} f_j(r_j, r_j^*, a, a^*)$$

where

$$f_j(r_j, r_j^*, a, a^*) = \begin{cases} \mathrm{pr}(R_j = 1 \mid A = a^*) & \text{if } r_j^* = r_j = 1, \\ \mathrm{pr}(R_j = 1 \mid A = a) - \mathrm{pr}(R_j = 1 \mid A = a^*) & \text{if } r_j^* = 0 \text{ and } r_j = 1, \\ 0 & \text{if } r_j^* = 1 \text{ and } r_j = 0, \\ \mathrm{pr}(R_j = 0 \mid A = a) & \text{if } r_j^* = r_j = 0. \end{cases}$$

Their second condition is no M–R additive mean interaction, i.e.,

$$E(Y \mid m, r, a) - E(Y \mid m^*, r, a) - E(Y \mid m, r^*, a) + E(Y \mid m^*, r^*, a) = 0,$$

for all levels m and m^* of M and r and r^* of R. For discrete M and R, this yields

$$\begin{aligned} \gamma_0 = & \sum_m \{E(Y \mid m, r^*, a) - E(Y \mid m^*, r^*, a)\} \,\mathrm{pr}(M = m \mid A = a^*) \\ & + \sum_r \{E(Y \mid m^*, r, a) - E(Y \mid m^*, r^*, a)\} \,\mathrm{pr}(R = r \mid A = a) \\ & + E(Y \mid m^*, r^*, a). \end{aligned}$$

Eschewing the cross-world-counterfactual assumptions of the NPSEM-IE , Tchetgen Tchetgen and Phiri (2014) extend the bounds of Robins and Richardson (2010) to allow for the presence of an exposure-induced confounder when the mediator is binary:

$$\begin{gathered} \max\left\{0, \mathrm{pr}(M = 0 \mid A = a^*) + \sum_r E(Y \mid M = 0, R = r, A = a)\mathrm{pr}(R = r \mid A = a) - 1\right\} \\ + \max\left\{0, \mathrm{pr}(M = 1 \mid A = a^*) + \sum_r E(Y \mid M = 1, R = r, A = a)\mathrm{pr}(R = r \mid A = a) - 1\right\} \\ \leq \gamma_0 \leq \\ \min\left\{\mathrm{pr}(M = 0 \mid A = a^*), \sum_r E(Y \mid M = 0, R = r, A = a)\mathrm{pr}(R = r \mid A = a)\right\} \\ + \min\left\{\mathrm{pr}(M = 1 \mid A = a^*), \sum_r E(Y \mid M = 1, R = r, A = a)\mathrm{pr}(R = r \mid A = a)\right\}. \end{gathered}$$

We extend these bounds as well to allow for polytomous M in Section 2.3. Additionally, we construct bounds for γ_0 under an NPSEM-IE that account for a discrete exposure-induced confounder, but require no further assumption.

2.3 New partial identification results

We begin by extending the bounds of Robins and Richardson (2010) and Tchetgen Tchetgen and Phiri (2014) to settings with discrete mediator and outcome.

Theorem 2.1. *Under a model corresponding to the* SWIG *in either Figure 2.1(b) or Figure 2.2(b) with discrete M and Y and arbitrary R,*

$$\begin{aligned}
&\sum_{m,y} y \left(\max\left[0, \mathrm{pr}\{M(a^*) = m\} + \mathrm{pr}\{Y(a,m) = y\} - 1\right] I(y > 0)\right. \\
&\qquad \left. + \min\left[\mathrm{pr}\{M(a^*) = m\}, \mathrm{pr}\{Y(a,m) = y\}\right] I(y < 0)\right) \\
&\qquad\qquad \leq \gamma_0 \leq \\
&\sum_{m,y} y \left(\max\left[0, \mathrm{pr}\{M(a^*) = m\} + \mathrm{pr}\{Y(a,m) = y\} - 1\right] I(y < 0)\right. \\
&\qquad \left. + \min\left[\mathrm{pr}\{M(a^*) = m\}, \mathrm{pr}\{Y(a,m) = y\}\right] I(y > 0)\right).
\end{aligned}$$

The upper and lower bounds coincide when $Y(a,m)$ or $M(a^*)$ is degenerate, which follows from the properties of joint probability mass functions. The upper bound is achieved only if $Y(a,m)$ and $M(a^*)$ are comonotone for each m, i.e., if $F_{Y(a,m),M(a^*)}(y,m) = \min\left[F_{Y(a,m)}(y), F_{M(a^*)}(m)\right]$ for each m; the lower bound is achieved only if they are countermonotone for each m, i.e., if $F_{Y(a,m),M(a^*)}(y,m) = \max\left\{0, F_{Y(a,m)}(y) + F_{M(a^*)}(m) - 1\right\}$ for each m. A straightforward application of the g-formula under the DAGs in Figures 2.1(a) and 2.2(a) yields the following corollaries:

Corollary 2.2. For polytomous M and Y, γ_0 is partially identified under the model corresponding to the SWIG in Figure 2.1(b) by the bounds in Theorem 2.1 with $\mathrm{pr}\{M(a^*) = m\} = \mathrm{pr}(M = m \mid a^*)$ and $\mathrm{pr}\{Y(a,m) = y\} = \mathrm{pr}(Y = y \mid m, a)$. It is partially identified under the model corresponding to the SWIG in Figure 2.2(b) by the same bounds, but with $\mathrm{pr}\{M(a^*) = m\} = \mathrm{pr}(M = m \mid a^*)$ and $\mathrm{pr}\{Y(a,m) = y\} = \sum_r \mathrm{pr}(Y = y \mid m, r, a)\mathrm{pr}(R = r \mid a)$.

The second part of the corollary continues to hold even if there were a hidden common cause of R and Y as in Figure 2.3, since the same g-formula applies in this setting. Whereas the previous results invoked no cross-world-counterfactual independences under the SWIG interpretation of the DAG in Figure 2.2(a), sharper bounds are available under Pearl's NPSEM-IE interpretation of these

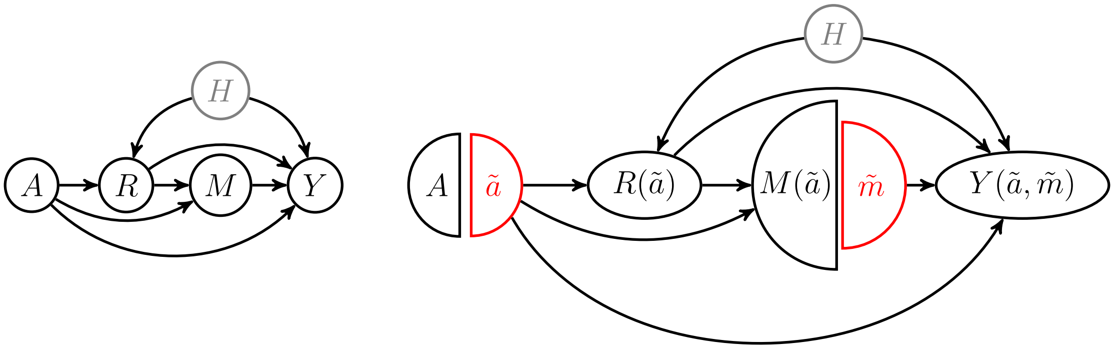


Figure 2.3: (a) A mediation directed acyclic graph in which an unobserved variable H affects R, an exposure-induced confounder, and Y. (b) The single-world intervention graph in the setting of (a) that has been intervened on to set A to $\tilde{a} \in \{a, a^*\}$ and M to $\tilde{m}$.

DAGs, as derived in the following result.

Theorem 2.3. *For discrete R taking values in $\{1, \ldots, p\}$, let B be the $p^2 \times (p-1)^2$ matrix*

$$\begin{bmatrix}
I_{p-1} & 0_{(p-1)\times(p-1)} & \cdots & 0_{(p-1)\times(p-1)} & 0_{(p-1)\times(p-1)} \\
-1^T_{p-1} & 0^T_{p-1} & \cdots & 0^T_{p-1} & 0^T_{p-1} \\
0_{(p-1)\times(p-1)} & I_{p-1} & \cdots & 0_{(p-1)\times(p-1)} & 0_{(p-1)\times(p-1)} \\
0^T_{p-1} & -1^T_{p-1} & \cdots & 0^T_{p-1} & 0^T_{p-1} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0_{(p-1)\times(p-1)} & 0_{(p-1)\times(p-1)} & \cdots & I_{p-1} & 0_{(p-1)\times(p-1)} \\
0^T_{p-1} & 0^T_{p-1} & \cdots & -1^T_{p-1} & 0^T_{p-1} \\
0_{(p-1)\times(p-1)} & 0_{(p-1)\times(p-1)} & \cdots & 0_{(p-1)\times(p-1)} & I_{p-1} \\
0^T_{p-1} & 0^T_{p-1} & \cdots & 0^T_{p-1} & -1^T_{p-1} \\
-I_{p-1} & -I_{p-1} & \cdots & -I_{p-1} & -I_{p-1} \\
\multicolumn{5}{c}{1^T_{(p-1)^2}}
\end{bmatrix},$$

d be the p^2-dimensional vector

$$\begin{bmatrix} 0_{p-1} \\ \mathrm{pr}\,(R=1 \mid A=a) \\ 0_{p-1} \\ \mathrm{pr}\,(R=2 \mid A=a) \\ \vdots \\ 0_{p-1} \\ \mathrm{pr}\,(R=p-1 \mid A=a) \\ \mathrm{pr}\,(R=1 \mid A=a^*) \\ \mathrm{pr}\,(R=2 \mid A=a^*) \\ \vdots \\ \mathrm{pr}\,(R=p-1 \mid A=a^*) \\ \mathrm{pr}\,(R=p \mid A=a) + \mathrm{pr}\,(R=p \mid A=a^*) - 1 \end{bmatrix},$$

and x be the vectorization of the matrix $\left[E\left\{E(Y \mid M, R=r, A=a) \mid R=r^*, A=a^*\right\}\right]_{r,r^*}$. Under a NPSEM-IE *corresponding to the* DAG *in Figure 2.2(a) where M and Y can be either continuous or discrete, γ_0 is partially identified by $\left[x^T(B\delta_L + d), x^T(B\delta_U + d)\right]$, where δ_L and δ_U are the minimizing and maximizing solutions respectively to the linear programming problem with objective function $x^T B\delta$ subject to the constraints*

$$\begin{bmatrix} I_{(p-1)^2} \\ -I_{(p-1)^2} \end{bmatrix} \delta \leq \begin{bmatrix} \min\{\mathrm{pr}(R=1 \mid A=a), \mathrm{pr}(R=1 \mid A=a^*)\} \\ \min\{\mathrm{pr}(R=1 \mid A=a), \mathrm{pr}(R=2 \mid A=a^*)\} \\ \vdots \\ \min\{\mathrm{pr}(R=p \mid A=a), \mathrm{pr}(R=p-1 \mid A=a^*)\} \\ \min\{\mathrm{pr}(R=p \mid A=a), \mathrm{pr}(R=p \mid A=a^*)\} \\ 1-\mathrm{pr}(R=1 \mid A=a) - \mathrm{pr}(R=1 \mid A=a^*) \\ 1-\mathrm{pr}(R=1 \mid A=a) - \mathrm{pr}(R=2 \mid A=a^*) \\ \vdots \\ 1-\mathrm{pr}(R=p \mid A=a) - \mathrm{pr}(R=p-1 \mid A=a^*) \\ 1-\mathrm{pr}(R=p \mid A=a) - \mathrm{pr}(R=p \mid A=a^*) \end{bmatrix}$$

and $\delta \geq 0$.

Similar to the previous result, these bounds coincide if either $R(a)$ or $R(a^*)$ is degenerate. The upper bound is achieved when $R(a)$ and $R(a^*)$ are comonotone; the lower bound is achieved when they are countermonotone. While these bounds are not available in closed form, they can be readily solved using standard software, such as with the lp_solve function, which uses the revised simplex method and is accessible from a number of languages, including R, MATLAB, Python, and C. While the method used by this software is not guaranteed to converge at a polynomial rate (Klee and Minty, 1970), it is quite efficient in most cases (Schrijver, 1998). The following corollary shows that these bounds reduce to a closed form when R is binary.

Corollary 2.4. Under a NPSEM-IE corresponding to the DAG in Figure 2.2(a) with binary R,

$$\min_{\pi_{11}\in\Pi}\sum_{r,r^*} E\left\{E(Y \mid M, R=r, A=a) \mid R=r^*, A=a^*\right\} h(r,r^*,\pi_{11})$$

$$\leq \gamma_0 \leq$$

$$\max_{\pi_{11}\in\Pi}\sum_{r,r^*} E\left\{E(Y \mid M, R=r, A=a) \mid R=r^*, A=a^*\right\} h(r,r^*,\pi_{11})$$

where Π is the set

$$\begin{aligned}&\{\max\{0, \mathrm{pr}(R=1 \mid A=a)+\mathrm{pr}(R=1 \mid A=a^*)-1\},\\ &\min\{\mathrm{pr}(R=1 \mid A=a), \mathrm{pr}(R=1 \mid A=a^*)\}\}\end{aligned}$$

and

$$h(r,r^*,\pi_{11}) = \begin{cases} \pi_{11} & \text{if } r^*=r=1,\\ \mathrm{pr}(R=1 \mid A=a)-\pi_{11} & \text{if } r^*=0 \text{ and } r=1,\\ \mathrm{pr}(R=1 \mid A=a^*)-\pi_{11} & \text{if } r^*=1 \text{ and } r=0,\\ 1-\mathrm{pr}(R=1 \mid A=a)-\mathrm{pr}(R=1 \mid A=a^*)+\pi_{11} & \text{if } r^*=r=0. \end{cases}$$

Under $A-R$ monotonicity with binary R, the identifying functional given by Tchetgen Tchetgen and VanderWeele (2012) is recovered at the upper bound in Corollary 2.4. All results given here can be extended to settings with observed pre-exposure confounders, which we denote C. In Corollary 2.2, one must first perform conditional inference given C, then subsequently average over the conditional bounds. This is in fact valid due to Jensen's inequality, because the constraints on the marginal joint probabilities are already implied by the constraints enforced on the conditional joint distributions, so no further constraints need be considered. Jensen's inequality does not apply in the case of Theorem 2.3, however, so controlling for C requires estimating two pairs of candidate bounds and selecting the larger of the lower bounds and the smaller of the upper bounds. When p is of moderate size, δ can be solved for each covariate pattern of C, i.e., without modeling the dependence of the cross-world-counterfactual joint distribution on C. Averaging the resulting conditional bounds gives the first pair of bounds. The second pair results from replacing each probability in the theorem with an average over the probabilities conditional on C and doing the same with x.

2.4 Application to Harvard PEPFAR data set

We now consider an application to a data set collected by the Harvard President's Emergency Plan for AIDS Relief (PEPFAR) program in Nigeria. The data set consists of previously antiretroviral therapy (ART)-naïve, HIV-1 infected adult patients who began ART in the program and were followed at least one year following initiation. Patients without reliable viral load data at two of the hospitals were excluded. Only complete cases initially assigned to either TDF+3TC/FTC+NVP or AZT+3TC+NVP[1] were considered for this analysis. Thus, the data set we consider consists of 6627 patients, 1919 of whom were assigned to TDF+3TC/FTC+NVP, and the remaining 4708 assigned to AZT+3TC+NVP.

There has accumulated evidence of a differential effect on virologic failure (defined by the World Health Organization as repeat viral load above 1000 copies/mL after 6 months of ART duration) between these two first-line antiretroviral treatment regimens (Tang et al., 2012). A natural question of scientific interest is what role adherence plays in mediating this differential effect. We are primarily interested in learning about the scientific mechanism of this effect on the individual level. The natural indirect effect best captures this mechanism in that it captures an isolated effect difference mediated by adherence by, in a sense, deactivating effect differences along all other possible causal pathways. We specifically examine the effect through adherence over the second six months since treatment assignment, i.e., the six months prior to the first viral load measurement. Identification is complicated by the presence of treatment toxicity, which clearly affects adherence directly, and has the potential to modify the effect of the treatment assignment on virologic failure. Thus, toxicity measured at six months after treatment assignment is an exposure-induced confounder of the effect of the mediator on the outcome. Further, toxicity and virologic failure are likely to be rendered dependent by unobserved underlying biological common causes as in Figure 2.3, where H represents these hidden biological mechanisms. Because we define the mediator to be adherence over the second six months, adherence over the first six months is also an

[1] 3TC=lamivudine, AZT=zidovudine, FTC=emtricitabine, NVP=nevirapine, TDF=tenofovir

exposure-induced confounder along with toxicity, and must be accounted for. Had we defined the mediator to be adherence over the full year, measurement of the mediator and toxicity would have overlapped, violating the principle of temporal ordering.

Let C denote the vector consisting of baseline covariates sex, age, marital status, WHO stage, hepatitis C virus, hepatitis B virus, CD4 count, and viral load. Let A be an indicator of ART assignment taking levels a^* for TDF+3TC/FTC+NVP and a for AZT+3TC+NVP; R be a vector of two indicator variables, one of the presence of any lab toxicity, and one of adherence exceeding 95%, both over the first six months following initiation of therapy; M be an indicator of adherence exceeding 95% over the subsequent six months; and Y be an indicator of virologic failure at one year, i.e., repeat viral load above 1000 copies/mL at one year and at 18 months.

Here we estimate the natural indirect effect of A on Y through M, as defined above, on the risk difference scale using the various sets of identifying assumptions given above. Throughout, inference is performed using maximum likelihood for point estimation and a randomly weighted bootstrap (Rao and Zhao, 1992; Van Der Vaart and Wellner, 1996) for confidence intervals due to the rarity of the outcome. The results are summarized in Figure 2.4. It is immediately apparent that inference is remarkably sensitive to which identifying assumptions are made. Consider an investigator who is willing to rely on cross-world-counterfactual independences. If she decides to ignore the presence of toxicity, she will conclude that there is a very small, yet significant negative indirect effect. Conversely, if she makes the no M–R interaction assumption, she will find a significant positive indirect effect with considerable uncertainty. In fact, an empirical test of this assumption reveals that it is unlikely to apply. Similarly, another empirical test suggests the assumption of independent errors of the components of R needed for the identifying functional under monotonicity to be valid is also unlikely to hold. Nonetheless, we present both results for the sake of comparison. Results are fairly imprecise under monotonicity, and do not show a significant effect.

Another investigator unwilling to impose cross-world-counterfactual-independence assumptions is left with little to say as the bounds are uninformative, regardless of how toxicity is handled. Interestingly, the bounds that result from making no assumptions about the joint distribution of the cross-world R counterfactuals are narrower than the bounds that result from ignoring R. That is, the bounds themselves appear narrower; the variances of the interval estimates appear to be com-

Figure 2.4: A plot showing the estimated natural indirect effect of ART assignment on virologic failure with respect to adherence under the various assumptions. The assumptions vary across the horizontal axis, with the first part of the label indicating the assumption regarding the exposure-induced confounder, R, and the second part indicating the assumption regarding cross-world counterfactuals. For the assumptions regarding R, “Ignore" means that the presence of R is ignored altogether, “Monoton." means the A–R monotonicity assumption in Section 2.1, “No M*R" means the no M–R interaction assumption in Section 2.1, and “None" means that R was accounted for without additional assumptions. For the assumptions regarding cross-world counterfactuals, “IE" means a NPSEM-IE was assumed, and “None" means no cross-world-counterfactuals independences were assumed. When the assumptions give partial identification, the two dots represent the point estimates of the upper and lower bound for the natural indirect effect, and the vertical bar represents the bootstrap 95% confidence interval for the interval. When the assumptions give full identification, the single dot represents the point estimate of the natural indirect effect, and the vertical bar represents its bootstrap 95% confidence interval.

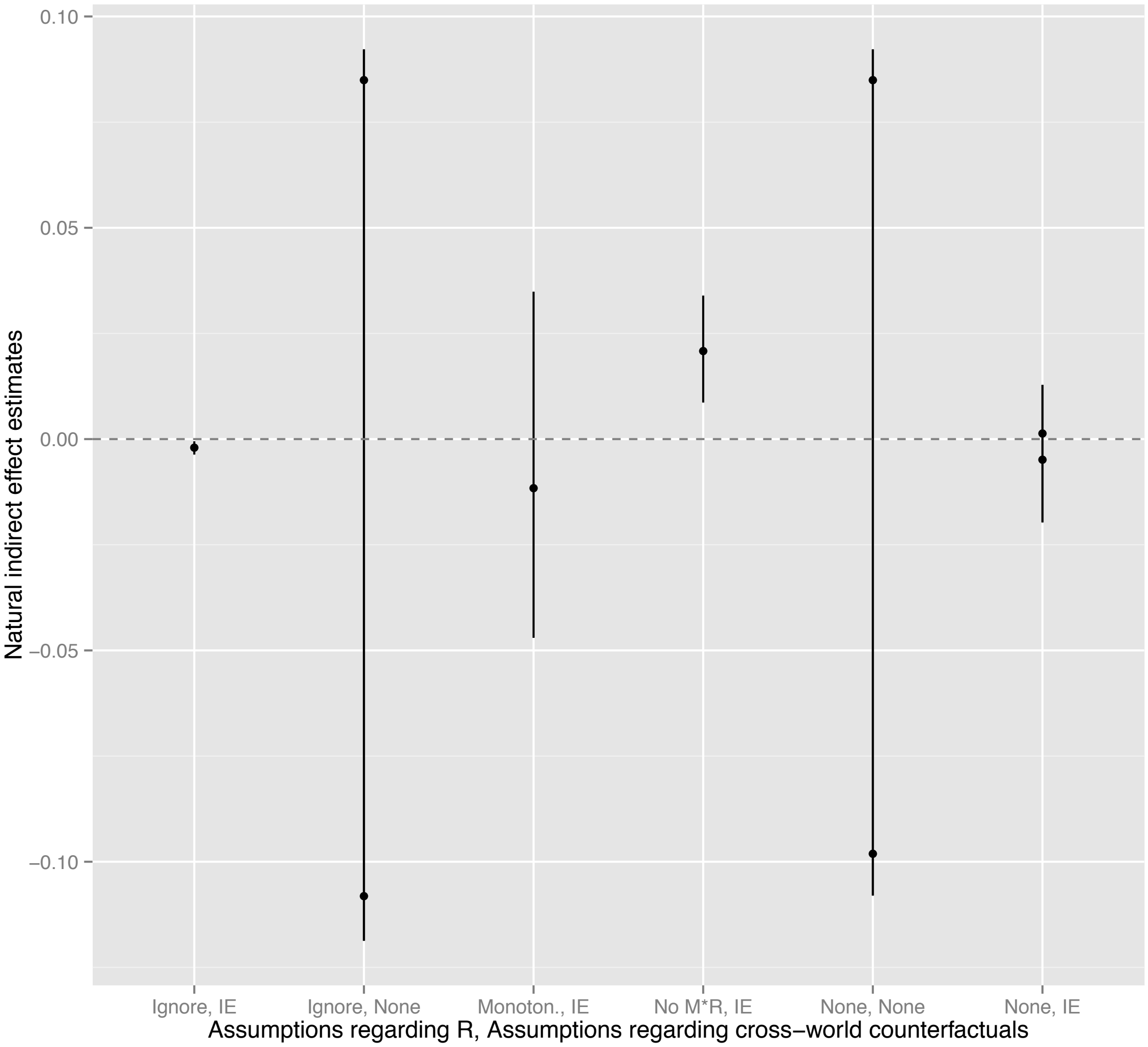


Figure 2.4: (Continued)

parable. This is because even though we do not impose any restrictions on the distribution of R or its counterfactuals a priori, observing R is clearly informative. The bounds accounting for R have the added advantage of being the only identifying formula that remains valid when toxicity and virologic suppression are affected by an unobserved common cause, as in Figure 2.3.

Finally, incorporating R results in narrower interval estimates than not imposing the independence assumption even if R were ignored. Thus, cross-world-counterfactual-independences appear to have stronger empirical implications in the current analysis than assumptions regarding exposure-induced confounders. The general trend in these results is that little is gained in terms of precision by assumptions regarding R. In fact, the confidence interval for the bounds resulting from the independent errors assumption and no assumption regarding R is narrower than the confidence interval for the estimate that results from assuming monotonicity, despite the fact that the NIE is point-identified in the latter case. The naïve assumption that R is not a confounder is the only assumption about R under which precision is gained.

A Class of Semiparametric Tests of Treatment Effect Robust to Measurement Error of a Confounder

Caleb H. Miles, Joel Schwartz, and Eric J. Tchetgen Tchetgen

Department of Biostatistics

Harvard School of Public Health

3.1 Introduction

In observational studies across an array of disciplines, it is not uncommon to observe variables X measured with error. As noted in Cote and Buckley (1987), “Campbell (1969) has gone so far as to say that measurement error (both random error and method effect) and its confounding influences on research findings cannot be avoided.” In the field of causal inference, data on covariates are needed to adjust for confounding in order to make inferences with causal interpretations. While a commonly-cited result states that under relatively-simple settings, the ordinary least squares (OLS) coefficient estimate of a single variable measured with error in a multiple linear regression will merely be attenuated to the null, and hence produce a valid (if conservative) hypothesis test of association, the effects of confounders measured with error can be more harmful. Unaccounted for, measurement error of confounders will produce biased effect estimates and invalid hypothesis tests in even the simplest of settings; a multiple linear regression of outcome Y on exposure A and confounders C and X – where the true confounder X^* is measured with classical, nondifferential error ε^* such that $X = X^* + \varepsilon^*$ – will produce a treatment effect estimate that is biased towards the crude (unadjusted) estimate. Consequently, any hypothesis test based on this regression will likely be invalid.

Though there has been tremendous interest in methodology accounting for confounders measured with error, previous research has largely focused on the case of misclassification of discrete confounders. In particular, Greenland (1980) conjectured that misclassification of a confounder results in a “partial” loss of control for confounding, in the sense that the average causal effect will be biased in the direction of the crude (unadjusted) estimate, but bound between this value and the truth. This was generally accepted as true (Fung and Howe, 1984; Kaldor and Clayton, 1985; Tzonou et al., 1985; Kelsey, 1996; Espeland and Hui, 1987) until shown by Ogburn and VanderWeele (2012) not to hold in general. The authors showed that this would hold in all-scalar-, binary-variable settings when there is no qualitative interaction between exposure and confounder. Ogburn and Vanderweele (2013) extended these conditions to settings with a polytomous confounder. Less attention has been given to the effect of measurement error on continuous confounders. Using parametric models, Cochran and Rubin (1973) gave an expression characterizing the bias incurred by such error that, under simplifying assumptions, allows for a correction provided the reliability ratio is known. Battistin and Chesher (2014) generalized this work to nonparametric models using a first-order approximation of bias for small measurement error (valid for a reliability ratio up to 70%) developed in Chesher (1991). This generalization allows for identification of the average causal treatment effect and the effect of treatment on the treated for a known measurement-error variance, or for sensitivity analysis by varying the measurement-error variance across a range of plausible values. Little else has been published since Cochran and Rubin (1973) on measurement error of a continuous confounder. Indeed, Battistin and Chesher (2014) claimed this work to be "to the best of [their] knowledge the only paper to study covariate measurement error within a programme evaluation [i.e., causal inference] context."

The method presented by Battistin and Chesher (2014) fits into a more general body of measurement error research that accounts for error without relying on external information. While most traditional measurement-error-adjustment methods depend on auxiliary data such as instrumental variables or data from reliability or validation studies, attention has more recently shifted to developing methods not depending on such data, which can be expensive to collect or simply unavailable. One such class of methods uses "higher-order" moment restrictions to produce identifying estimating equations for parameters from the deterministic component of a regression model with covariates measured with error. Such identifying moment restrictions are implied by the assumption of mutual independence between the error-free covariates, the measurement error, and the residuals, which can often be justifiable. The setting in which the regression model is linear in its covariates has been well studied (Kapteyn and Wansbeek, 1983; Stuart and Kendall, 1979; Pal, 1980; Cragg, 1997; Dagenais and Dagenais, 1997; Lewbel, 1997; Erickson and Whited, 2000, 2002; Lewbel, 2012; Bonhomme and Robin, 2009, among others), and has been shown to be identified under simple conditions. Schennach and Hu (2013) showed that under sufficient regularity conditions, nonlinear regression models are nonparametrically identified outside of a particular parametric class depending on four parameters.

Another method free of external information, known as deconvolution, couples knowledge of the measurement-error distribution and a kernel estimator of an error-ridden variable to, in a sense, "cancel out" the measurement error and recover the density of the error-free variable. Fan (1991) showed that under certain smoothness conditions, convolution density estimates will have convergence rates of $n^{-\eta}$ for some $\eta > 0$. Fan and Truong (1993) extended this technique to nonparametric regression. Despite this, convergence rates often tend to be too slow for practical use. Furthermore, though the idea of convolution is attractive in principle, it is very rare that the distribution of measurement error will be known in practice.

When conditions for point identification are not met, it can be possible to compute bounds within which the parameter of interest is partially identified. Frisch (1934) first explored this subject, giving bounds in the simple linear regression case, which Klepper and Leamer (1984) later extended to multivariable linear regression with all covariates measured with error. Set identification is more challenging in nonlinear regression models, in that it is not available in closed form. Schennach (2014) proposed a simulation-based method (ELVIS) for partial identification for a class of latent-variable models that contains the nonlinear regression model with covariate measurement error as a special case. Carroll et al. (2006) and Schennach (2012) give surveys of measurement error methodology containing a more thorough treatment of methods not requiring external information.

This paper contributes to both the literature on confounder measurement error as well as on measurement error methods not requiring external information. We present a large class of valid test statistics of the null hypothesis of no average causal effect when a set of continuous confounders are measured with classical, nondifferential error. This class of test statistics is of interest in a variety of practical settings in that they require neither knowledge of the distribution or variance of the measurement error (as in Battistin and Chesher (2014) or in deconvolution), nor any form of external information. We will assume that the measurement error is independent of the error-free confounders and all other observed variables. Otherwise, no other moment restrictions are required that are not already embedded in the assumptions needed to draw causal inferences in the absence of measurement error. In particular, we leverage the no-unobserved-confounding assumptions needed for identification of the average causal effect.

The governing idea of the proposed approach is that under the null hypothesis of no effect of exposure, the assumption of no unobserved confounding renders the outcome an instrumental variable (IV) for the association between the true error-prone covariate X^* and A adjusting for C. Thus, as documented in the literature on IV methods for measurement error, Y can be used to obtain a consistent estimator of the association between (C, X^*) and A (Amemiya, 1985; Amemiya et al., 1990; Amemiya, 1990; Carroll and Stefanski, 1994; Stefanski and Buzas, 1995; Buzas and

Stefanski, 1996; Carroll et al., 2006; Fuller, 2009, among others). In this paper, we show that estimation of the conditional association between the error-prone covariates and exposure can be accomplished jointly with a test of no treatment effect under a unifying framework of a generalized method of moments test based on overidentifying moment restrictions, known in the econometrics literature as a Sargan test (Sargan, 1958), Hansen test, or J-test (Hansen, 1982).

Our class of test statistics includes semiparametric tests with nominal type I error rate, if under the null hypothesis of no treatment effect, (1) the conditional association between the mismeasured covariates and exposure is linear (on the additive, multiplicative or logit scale) given error-free confounders, and either (2.a) the association between error-free confounders and exposure given the true error-prone confounders is correctly specified or (2.b) the association between the outcome and error-free covariates is correctly specified. Included in this class are tests that are doubly robust in the sense that they preserve nominal type I error rate if at least one of (2.a) or (2.b) holds, and it need not be known which conditions holds. Validity of the robust propensity-score tests and doubly-robust tests is demonstrated via simulation studies. We conclude with a data application in which we test for a causal effect of same-day temperature on mortality in the United States. We conduct a multi-city analysis with daily information on mortality as well as environmental factors including temperature and concentration of particulate matter with diameter of 2.5 micrometres or less (PM2.5). PM2.5 is known to be a confounder and to be measured with error due to the high level of variability of pollution across monitoring stations (Armstrong, 1990; Zeger et al., 2000; Armstrong, 2004; Bateson et al., 2007; Kioumourtzoglou et al., 2014). We apply our method to test for this effect, controlling for seasonal trends and PM2.5 for confounding while accounting for the measurement error of the latter.

3.2 A class of propensity-score-based test statistics robust to measurement error

To formalize discussion, we define for each a the counterfactual Y_a to be a subject's outcome had the subject been assigned, possibly contrary to fact, to exposure level a. We link these counterfactuals to the observed variables via the consistency assumption (Robins, 1986), which states that if $A = a$, then $Y_a = Y$ with probability one for each level a. We assume the observed covariates X

are measured with classical error, i.e., additive independent measurement error, and are related to their corresponding, unobserved, true value X^* by $X = X^* + \varepsilon^*$, where ε^* is the measurement error, assumed to be independent of X^*, C, A, and Y. Suppose that in addition to X^*, a set of error-free confounders C are also observed. Further, suppose the following holds:

Assumption 3.1. *$Y_a \perp\!\!\!\perp A \mid C, X^*$ for each level a (No unmeasured confounding).*

In the following developments, suppose that A is continuous; results are generalized to binary and count exposure in Section 3.8. We now present a class of test statistics for the null hypothesis

$$H_0 : E(Y_a \mid C, X^*) = E(Y_0 \mid C, X^*) \text{ for all } a.$$

Intuitively, under the sharp null, the assumption of no unmeasured confounding implies that

$$A \perp\!\!\!\perp Y \mid C, X^*. \tag{3.1}$$

Furthermore since X^* is a confounder, we have that

$$Y \not\perp\!\!\!\perp X^* \mid C. \tag{3.2}$$

Statements 3.1 and 3.2 formally define Y as an instrumental variable for the conditional association between X^* and A given C (Carroll et al., 2006). Although H_0 is weaker than the sharp null, this result continues to hold despite Y no longer formally being an IV.

We present our first result, which relies on correct specification of a mean regression model for exposure. In this vein, consider the semiparametric model $\mathcal{M}_A$ as the set of laws for (A, X^*, C, Y) with sole restriction the parametric model

$$E(A \mid C, X^* = 0) = g_A(C; \alpha_1)$$
$$E(A \mid C, X^*) - E(A \mid C, X^* = 0) = \alpha_2^T X^*,$$

where g is a known function of C indexed by α_1. The unknown parameters α_1 and α_2 have dimensions p_1 and p_2, respectively. We assume throughout that the $X^* - A$ association is linear given C, however $g_A(C; \alpha_1)$, though parametric, can be any nonlinear function in C.

Theorem 3.2. *Let $\ell(C)$ and $m(C)$ each be vector-valued functions of C with linearly-independent* *elements and dimension $p + q$ for some positive integer q, where $p = p_1 + p_2$, and let*

$$U(\alpha) \equiv \{\ell(C)Y + m(C)\} \{A - E(A \mid C, X; \alpha)\},$$

where $E(A \mid C, X; \alpha) = g_A(C; \alpha_1) + \alpha_2^T X$. Further, let $\Omega = E\left[U(\alpha)U(\alpha)^T\right]$ and $\hat{U}_n(\alpha) \equiv \mathbb{P}_n U_i(\alpha)$, where $\mathbb{P}_n$ denotes the empirical mean of its argument. If $U(\alpha)$ is continuously differentiable, $\nabla_\alpha E\{U(\alpha)\} = E\{\nabla_\alpha U(\alpha)\}$, and $\Omega^{-1}E\{\nabla_\alpha U(\alpha)\}$ has full rank, then for any $\hat{\Omega}_n \xrightarrow{p} \Omega$, the test statistic

$$\chi^2_{\text{robust}_A} \equiv \min_\alpha n\hat{U}_n(\alpha)^T \hat{\Omega}_n^{-1} \hat{U}_n(\alpha) \xrightarrow{d} \chi^2_q$$

under $\mathcal{M}_A$ and H_0.

Thus, we have a valid test which depends on X^* only through the mismeasured covariate, X. Intuitively, the standard normal equations for coefficients in the propensity-score model incur bias due to the components that are the product of the residual with the error-prone covariate. However, this can be amended by replacing these components with the product of the residual with Y, since under H$_0$, this product forms an unbiased estimating equation. More such unbiased estimating equations can be added simply by multiplying his product by functions of C, and hence these can be used to form a valid Sargan test statistic. Thus, a simple form of the test in Theorem 3.2 with $q = 1$ could use $\nabla_{\alpha_1} g_A(C; \alpha)$ augmented by Y and the product of Y with an element of C in place of $\ell(C)Y + m(C)$, for instance. The following iterative procedure can be used to compute the variance-estimate component $\hat{\Omega}_n$:

```
initialize $\tilde{\alpha} := \arg\min_\alpha \hat{U}_n(\alpha)^T \hat{U}_n(\alpha)$;
set $\hat{\Omega}_n := \mathbb{P}_n\left[U(\tilde{\alpha})U(\tilde{\alpha})^T\right]$;
do
    set $\tilde{\alpha} := \arg\min_\alpha \hat{U}_n(\alpha)^T \hat{\Omega}_n^{-1} \hat{U}_n(\alpha)$;
    set $\hat{\Omega}_n := \mathbb{P}_n\left[U(\tilde{\alpha})U(\tilde{\alpha})^T\right]$;
while convergence not reached;
```

The first two steps are in fact sufficient for asymptotic validity, however iterating generally improves finite-sample performance. Alternatively, a continuous updating approach can be used, in which $\hat{\Omega}_n$ is indexed by α, and $n\hat{U}_n(\alpha)^T \hat{\Omega}_n(\alpha)^{-1} \hat{U}_n(\alpha)$ is minimized in α through both $\hat{U}_n(\alpha)$ and $\hat{\Omega}_n(\alpha)$. Next we illustrate the behavior of the proposed test statistic via a simulation study.

3.3 A simulation study of the test statistic $\chi^2_{\text{robust}_A}$

We now present hypothesis testing results from a simulation study drawing samples from the following data generating mechanism. We generate (Y_0, C) under a joint normal model given by

$$\begin{aligned} Y_0 &= N(0,1) \\ C &= Y_0 + N(0,1), \end{aligned}$$

and X^* and A under

$$\begin{aligned} X^* &= Y_0 + C + Y_0 C + N(0,1) \\ A &= C + X^* + N(0,4). \end{aligned}$$

To reflect the sharp null hypothesis, we let $Y = Y_0$. We generate X from the classical measurement error model

$$X = X^* + N(0, 9[1/\tau - 1]),$$

where τ is the reliability ratio, i.e., the ratio of the variability of the correctly-measured variable X^* to the variable measured with error X. One may easily verify that Assumption 3.1 and H_0 are satisfied.

Samples were repeatedly drawn under twelve different settings: with sample sizes of 1000, 5000, and 10,000, each with reliability ratios of 50%, 70%, 90%, and 100% (i.e., no measurement error). In each setting, we applied the simple form of the testing procedure described in Section 3.2, using a Gram-Schmidt orthonormalization of $[1, C, C^2, C^3]$ for the function $\ell(C)$, $m(C) = 0$, $g_A(C; \alpha_1) = \alpha_1 C$, and $q = 1$. We compared this test with two others that ignored the presence of measurement error. The first was a robust outcome-regression test, using the p-value of the regression coefficient for A when regressing Y on C, X, and A and using a sandwich variance estimate. Though this model is not correctly specified, the OLS estimate of the coefficient for A in the absence of measurement error will be unbiased for the slope of A in $E(Y \mid A, C, X^*)$ (zero). This is because it is equal to the OLS estimate of the coefficient for the residual obtained from a linear regression of A on C and X^*, which happens to be correctly specified. The second comparison test was a g-estimation test (Robins, 1989), using the p-value of the regression coefficient for Y when regressing A on C, X, and Y. All tests used an α level of 0.05. To demonstrate the validity of our test, we conducted all tests on 100,000 samples per setting. The results are presented in Table 3.1.

Table 3.1: Estimated type I error rate from 100,000 hypothesis tests simulated under the null hypothesis

n	Rel. ratio (%)	Robust PS	G-estimation	Standard OR
1000	50	0.0604	1.000	1.000
	70	0.0444	0.989	0.985
	90	0.0431	0.493	0.485
	100	0.0447	0.0511	0.0508
5000	50	0.0485	1	1
	70	0.0482	1	1
	90	0.0485	0.991	0.991
	100	0.0491	0.0496	0.0494
10,000	50	0.0493	1	1
	70	0.0495	1	1
	90	0.0501	1.000	1.000
	100	0.0497	0.0500	0.0498

This simulation clearly demonstrates that the standard OLS and g-estimation tests had approximately-correct type I error rates only in the absence of measurement error. In contrast, the proposed test statistic was found to have correct type I error rate across all settings considered, with improved performance with increasing sample size. There was no discernable trend between estimated type I error rates and reliability ratios for the robust test.

3.4 A class of doubly-robust test statistics

Validity of the test given in Section 3.2 relies on correct specification of $g_A(C;\alpha_1)$, however this model may be misspecified, thus it is of interest to explore an alternative, potentially more robust approach. Here we present a large class of doubly-robust test statistics. In order to describe this class of statistics, let $E(Y \mid C;\gamma) = g_Y(C;\gamma)$, and consider the semiparametric model $\mathcal{M}_Y$ with sole restrictions

$$E(A \mid C, X^*) - E(A \mid C, X^* = 0) = \alpha_2^T X^*$$

$$E(Y \mid C) = g_Y(C;\gamma).$$

This is a semiparametric model since the association between C and A given $X^* = 0$ is unrestricted. Further consider the union model $\mathcal{M}_\cup \equiv \mathcal{M}_A \cup \mathcal{M}_Y$. We present a class of test statistics for each of these two models, adopting the notation $\Delta_A(\alpha) \equiv A - E(A \mid C, X; \alpha)$ and $\Delta_Y(\gamma) \equiv Y - E(Y \mid C; \gamma)$ for the residuals in each model.

Theorem 3.3. *Let*

$$U(\alpha, \gamma) \equiv \begin{bmatrix} k(C)\Delta_Y(\gamma)(A - \alpha_2^T X) \\ S(\gamma) \end{bmatrix},$$

where $S(\gamma)$ is a system of estimating equations for γ, $k(C)$ is a vector-valued function of C with linearly-independent elements with dimension $p_2 + q$ for some positive integer q. Suppose $U(\alpha, \gamma)$ is continuously differentiable, $\nabla_{\alpha_2,\gamma} E\{U(\alpha, \gamma)\} = E\{\nabla_{\alpha_2,\gamma} U(\alpha, \gamma)\}$, and $\Omega^{-1} E\{\nabla_{\alpha_2,\gamma} U(\alpha, \gamma)\}$ has full rank, where $\Omega = E\left[U(\alpha, \gamma) U(\alpha, \gamma)^T\right]$. Then under $\mathcal{M}_Y$ and H_0, $\chi^2_{\text{robust}_Y} \equiv \min\limits_{\alpha_2,\gamma} n\hat{U}_n(\alpha, \gamma)^T \hat{\Omega}_n^{-1} \hat{U}_n(\alpha, \gamma) \xrightarrow{d} \chi^2_q$, for any $\hat{\Omega}_n \xrightarrow{p} \Omega$.

Theorem 3.4. *Let*

$$U(\alpha, \gamma) \equiv \begin{bmatrix} k(C)\Delta_Y(\gamma)\Delta_A(\alpha) \\ \{\ell(C)Y + m(C)\}\,\Delta_A(\alpha) \\ S(\gamma) \end{bmatrix},$$

where $S(\gamma)$ is a system of estimating equations for γ that is unbiased when $g_Y(C; \alpha)$ is correctly specified, $k(C)$, $\ell(C)$, and $m(C)$ are each vector-valued functions of C with linearly-independent elements, and ℓ and m have dimension p_1 and k has dimension $p_2 + q$ for some positive integer q. Suppose $U(\alpha, \gamma)$ is continuously differentiable, $\nabla_{\alpha,\gamma} E\{U(\alpha, \gamma)\} = E\{\nabla_{\alpha,\gamma} U(\alpha, \gamma)\}$, and $\Omega^{-1} E\{\nabla_{\alpha,\gamma} U(\alpha, \gamma)\}$ has full rank, where $\Omega = E\left[U(\alpha, \gamma) U(\alpha, \gamma)^T\right]$. Then under $\mathcal{M}_\cup$ and H_0, $\chi^2_{dr} \equiv \min\limits_{\alpha,\gamma} n\hat{U}_n(\alpha, \gamma)^T \hat{\Omega}_n^{-1} \hat{U}_n(\alpha, \gamma) \xrightarrow{d} \chi^2_q$, for any $\hat{\Omega}_n \xrightarrow{p} \Omega$.

As before, an appropriate variance estimator $\hat{\Omega}_n$ can be computed using either an iterated procedure or a continuous-updating approach. An alternative approach would be to first estimate γ by solving $\mathbb{P}_n S(\gamma) = 0$, plug this value into $U(\alpha, \gamma)$ (rendering the γ-estimating-equation component zero), and use

$$\hat{\Omega}_n = \frac{1}{n} \sum_{i=1}^{n} \left[U_i(\hat{\alpha}, \hat{\gamma}) - \frac{1}{n} \sum_{j=1}^{n} \nabla_\gamma U_j(\hat{\alpha}, \gamma) \mid_{\hat{\gamma}} \left\{ \frac{1}{n} \sum_{j=1}^{n} \nabla_\gamma S_j(\hat{\gamma}) \right\}^{-1} S_i(\hat{\gamma}) \right]^{\otimes 2}$$

for the variance estimator in the denominator of the test statistic.

All estimates of $\hat{\Omega}_n$ discussed so far require that k, ℓ, and m have not been estimated. Power for both these and the previous tests can be optimized by using appropriate choices of the functions $\ell(C)$, $m(C)$, and $k(C)$ based on the direction of the alternative hypothesis, which we discuss further in Section 3.7.

3.5 A simulation study for double robustness

We now demonstrate the double-robustness property via a simulation study. We drew 100,000 samples of size 5000 from the same data generating mechanism as in Section 3.3, and tested the same null hypothesis that $E(Y_a \mid C, X^*) = E(Y_0 \mid C, X^*)$ for all a using the doubly-robust form of our test described in the previous section. For comparison, we also include all tests considered in the previous simulation study of Section 3.3. All tests with the exception of the standard outcome-regression test (see Table 3.2 footnote) were conducted under three different models: the intersection model $\mathcal{M}_\cap$, in which both $g_Y(C;\gamma)$ and $g_A(C;\alpha_1)$ were correctly specified; $\mathcal{M}_Y$, in which $g_Y(C;\gamma)$ was correctly specified and $g_A(C;\alpha_1)$ was not; and $\mathcal{M}_A$, in which $g_A(C;\alpha_1)$ was correctly specified and $g_Y(C;\gamma)$ was not. We used $g_Y(C;\gamma) = \gamma_0 + \gamma_1 C$ and $g_A(C;\alpha_1) = \alpha_1 C$ for the

Table 3.2: Estimated type 1 error from 100,000 hypothesis tests simulated under the null hypothesis

Model	Rel. ratio (%)	DR	Robust PS	G-estimation	Standard OR[2]
$\mathcal{M}_\cap$	50	0.0455	0.0472	1	1
	70	0.0453	0.0482	1	1.000
	90	0.0533	0.0484	0.645	0.643
	100	0.0476	0.0489	0.0496	0.0493
$\mathcal{M}_Y$	50	0.0452	0.714	1	1
	70	0.0486	0.858	1	1.000
	90	0.0527	0.939	1	0.643
	100	0.0519	0.958	0.965	0.0493
$\mathcal{M}_A$	50	0.0463	0.0472	1	1
	70	0.0495	0.0482	1	1.000
	90	0.0472	0.0484	0.645	0.643
	100	0.0497	0.0489	0.0496	0.0493

[2] The conditional mean of Y given C and A does not have a simple form, yet the standard outcome-regression test does not require it to be modeled correctly for validity as described in Section 3.3. Therefore, we show results for the standard outcome-regression test using the misspecified model described in Section 3.3 in all cases for the purposes of comparison.

correctly-specified models and $g_Y(C;\gamma) = \gamma_0 + \gamma_1 C^2$ and $g_A(C;\alpha_1) = \alpha_1 C^2$ for the incorrectly-specified models. The index functions for the doubly-robust test used a Gram-Schmidt orthonormalization of $[1, C, C^2, C^3]$, with $k(C)^T$ and $m(C)^T$ being equal to the first two columns and $\ell(C)^T$ being equal to the last two. The score equations for γ were $S(\gamma) = [1, C^2]^T(Y - \gamma_0 - \gamma_1 C^2)$ under $\mathcal{M}_A$, and $S(\gamma) = [1, C]^T(Y - \gamma_0 - \gamma_1 C)$ otherwise. The results are presented in Table 3.2.

As expected, the doubly-robust test was approximately valid with correct Monte Carlo type 1 error rate under all settings. The standard test, on the other hand, was approximately valid under $\mathcal{M}_\cap$ and $\mathcal{M}_A$, but not under $\mathcal{M}_Y$. As in the previous study, the g-estimation and standard outcome-regression tests were not valid in the presence of measurement error, and the standard outcome-regression test was approximately valid in its absence. The g-estimation test was approximately valid under no measurement error only when g_A was correctly specified.

3.6 Application to test for an effect of temperature on mortality

As evidence for climate change continues to accumulate, the natural question of whether temperature affects mortality is of increasing public-health concern. While there are many long-term threats posed by rising global temperatures, the immediate effects on mortality also pose a grave public health concern. When studying this effect, it is vital to control for pollution as a potential confounder (O'Neill et al., 2003). A common metric of pollution is PM2.5 concentration, however this is well known to be measured with error (Armstrong, 1990; Zeger et al., 2000; Armstrong, 2004; Bateson et al., 2007; Kioumourtzoglou et al., 2014). In particular, PM2.5 is considered to be contaminated with a mixture of both Berkson error, due to the variability of concentration actually experienced across individuals, and classical error, due to aggregation of measurements across multiple monitoring stations (Zeger et al., 2000; Kioumourtzoglou et al., 2014). The former is benign in the sense that it increases variance but introduces no bias; it is the latter with which we are most concerned. Some studies have attempted to account for measurement error on PM2.5 using spatial smoothing models (Hoek et al., 2002; Jerrett et al., 2005; Yanosky et al., 2008; Puett et al., 2009; Szpiro et al., 2010; Sampson et al., 2011), however these rely on geographical data on residency and may induce other forms of error (Gryparis et al., 2009; Szpiro et al., 2011; Sheppard et al., 2012). We implemented our method, which does not depend on any collection of extraneous data, and compared it against two methods that ignore the presence of measurement error.

The data set used here consists of time-series mortality data from forty-one U.S. cities measured over the course of 1999 to 2006, though in our analysis, we only considered twenty-four cities with at least eight deaths per day, as cities with lower mortality rates were unlikely to provide enough power to detect an effect. Data on individual mortality with exact date of death was acquired from the National Center for Health Statistics (NCHS) and from state public health departments (Zanobetti et al., 2009). We excluded accidental deaths (ICD-code 10th revision: V01-Y98, ICD-code 9th revision: 1-799) and deaths of individuals who did not reside in the city in which they died. Temperature data were obtained from the National Oceanic and Atmospheric Administration (NOAA) website, with a city being assigned ambient temperature readings from its nearest monitoring station. PM2.5 data were obtained from the US Environmental Protection Agency's (EPA) Air Quality System (AQS) database (US EPA 2013). PM2.5 readings were averaged over all monitors in a city whenever multiple were available.

On a given day, i, let Y_i denote the number of deaths, A_i denote the average temperature in degrees Celcius, X_i denote the average PM2.5 concentration measurement, and C_i consist of date, t_i, and dummy variables for day of week. Models $g_A(C; \alpha)$ and $g_Y(C; \gamma)$ in the propensity-score and outcome-regression models used both Fourier bases for time with a period of one year to account for seasonal trends as well as polynomial bases for time to account for secular trends. The dimensions of the Fourier bases were at least four (not including intercept) and the dimensions of the polynomial bases started at zero. Sargan goodness-of-fit tests were used to assess model fit, and more dimensions were added to the bases until the tests no longer rejected at an α level of 0.10. In particular, we used forms of the test in Theorem 3.2 with $\ell(C) = 0$ (eliminating its power to test for an effect on Y) and $m(C)$ equal to $\nabla_\alpha g_A(C; \alpha)$ augmented by the next two polynomial or Fourier basis functions. Analogous tests were used for the robust outcome-regression model, with the moment functions being equal to the regression residuals multiplied by $\nabla_\gamma g_Y(C; \gamma)$ augmented by the next two polynomial or Fourier basis functions. Without this step, test rejections could be attributable to model misspecification rather than the presence of a true effect. Both models were linear in these terms as well as the day-of-week dummy variables. The robust outcome-regression model used a log link.

Measurement-error-robust tests from each of the three classes presented in Theorems 3.2-3.4

were conducted based on these models. For the robust propensity-score test, we used $[1, t, 0_{p-1}^T]^T$ as the function $\ell(C)$, where 0_{p-1} is a vector of zeroes with length $p-1$, and $[0, 0, \nabla_\alpha^T g_A(C;\alpha)]^T$ as the function $m(C)$. For the doubly-robust test, we used $[1, t]^T$ as the function $k(C)$, $[1, \nabla_\alpha^T g_A(C;\alpha)]^T$ as $m(C)$, and a vector of zeroes with length $p_1 + 1$ as the function $\ell(C)$. The same index function $k(C)$ was used for the robust outcome-regression test. Thus, in each case $q = 1$, and we compared resulting test statistics with the corresponding null distribution, χ_1^2. We used the doubly-robust test for inference, and supplemented our analysis with the other two for an additional check of model fit. The two standard statistics – based on the g-estimation and standard outcome-regression tests – were implemented as in the simulation studies, and used the same $g_A(C;\alpha)$ and $g_Y(C;\gamma)$ described above for the robust test statistics, with the addition of a linear term for temperature in the outcome-regression model. We used OLS to estimate the exposure model for the g-estimation test, a quasi-Poisson model for the standard outcome-regression test, and sandwich variance estimators for both.

The doubly-robust test rejected the null hypothesis of no effect of temperature on mortality in two cities: New York, NY and Stamford, CT. In each case, the robust propensity-score test also rejected and the robust outcome-regression test did not. In no city did the robust outcome-regression test reject, suggesting the possibility that it was underpowered. There was no indication of substantial attenuation due to measurement error for New York, as both standard tests rejected. For Stamford, on the other hand, it does appear that measurement error may have masked the effect from the standard tests, as they both failed to reject.

The robust propensity-score test rejected for both Allentown, PA and Philadelphia, PA, while neither the doubly-robust nor the robust outcome-regression test did, suggesting that the propensity-score model may not be correctly specified in these cases. Neither of the standard tests rejected in either of these cities as well, further reducing the credibility of the robust propensity-score tests' results for these two cities. In Newark, Salt Lake City, and Washington, one or more of the standard tests rejected when the robust tests did not. In these cases, we conjecture that the standard tests were invalidated by the presence of measurement error, and had we not accounted for it, we would have falsely claimed evidence for an effect in these cities when in fact there was none.

Table 3.3: P-values of hypothesis tests for an effect of temperature on mortality in U.S. cities

City	Robust PS	Robust OR	Doubly robust	G-estimation	Stand. OR
Albuquerque, NM	0.19	0.22	0.20	0.26	0.27
Allentown, PA	0.00096	0.72	0.92	0.26	0.21
Annandale, VA	0.60	0.34	0.79	0.95	0.96
Baltimore, MD	0.93	0.28	0.81	0.95	0.81
Boston, MA	0.39	0.93	0.37	0.051	0.054
Elizabeth, NJ	0.23	0.94	0.58	0.15	0.22
Hartford, CT	0.060	0.25	0.053	0.18	0.17
Lancaster, PA	0.18	0.54	0.17	0.20	0.32
Melville, NY	0.61	0.90	0.18	0.55	0.54
Middlesex, NJ	0.40	0.31	0.18	0.41	0.40
New Haven, CT	0.50	0.24	0.89	0.086	0.12
New York, NY	0.024	0.75	0.000059	0.0035	0.0017
Newark, NJ	0.61	0.41	0.53	0.038	0.053
Paterson, NJ	0.54	0.27	0.30	0.87	0.71
Philadelphia, PA	0.00000038	0.31	0.12	0.49	0.054
Reading, PA	0.52	0.11	0.64	0.29	0.28
Richmond, VA	0.19	0.44	0.28	0.98	0.98
Salt Lake City, UT	0.22	0.43	0.83	0.0048	0.0031
Spokane, WA	0.67	0.68	0.25	0.95	0.96
Stamford, CT	0.044	0.82	0.027	0.79	0.71
Upper Marlboro, MD	0.055	0.37	0.82	0.17	0.079
Washington, DC	0.92	0.85	0.80	0.011	0.014
Wilmington, DE	0.56	0.36	0.87	0.55	0.48
York, PA	0.25	0.83	0.26	0.34	0.35

3.7 Power and estimation under an additive causal model

In this section, we consider estimation and testing under the alternative hypothesis of an additive causal model,

$$E[Y_0 \mid A, C, X^*] = E[Y - \psi_0 A \mid A, C, X^*], \tag{3.3}$$

where ψ_0 is the average causal effect, i.e., the causal parameter of interest, though the following discussion can be easily adapted to other structural mean models. We conducted a supplementary simulation study varying the value of ψ_0 to demonstrate the local power of our proposed test statistics, using the same data generating mechanism as in Section 3.3 in each of the same measurement error settings, but with $Y = Y_0 + \psi_0 A$ to encode the alternative hypothesis. The robust propensity-score test was conducted on 1000 samples of size 5000 for each value of ψ_0. Results are presented in Figure 3.1. Additional results for sample sizes of 1000 and 10,000 can be found in the appendix. As expected, we observed trends of increasing power with effect size and reliability ratio. Also as expected, power was approximately 0.05 for $\psi_0 = 0$ in all cases. The test achieved an estimated 80% power at about $\psi_0 = 0.06$ when $\tau = 0.5$, about $\psi_0 = 0.05$ when $\tau = 0.7$, about $\psi_0 = 0.025$ when $\tau = 0.9$, and about $\psi_0 = 0.02$ when $\tau = 1$. The test's power appeared to be tending towards unity in all cases.

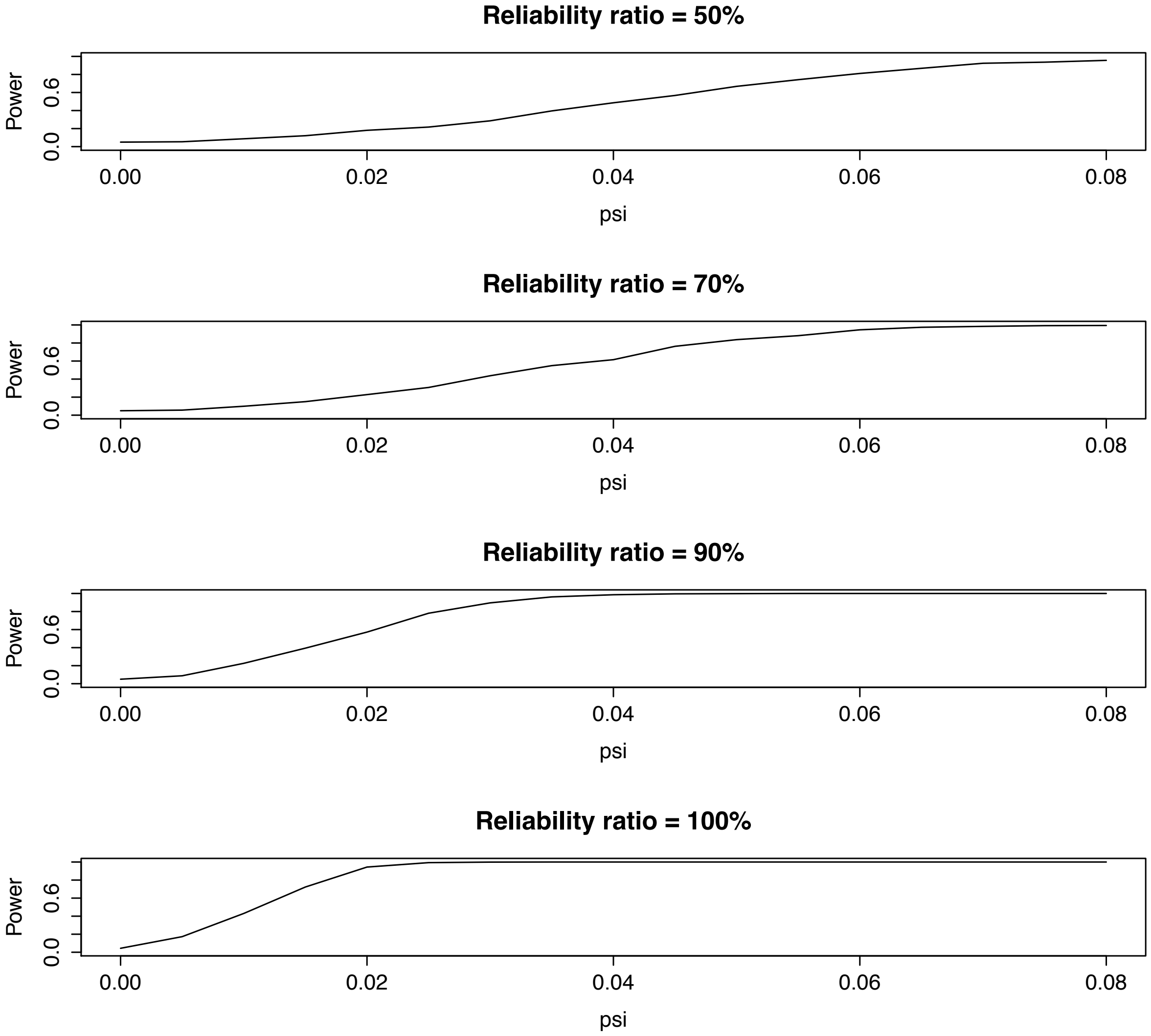


Figure 3.1: Simulation results demonstrating power for n=5000.

Having posited a model for the effect of A on Y, our testing approach can be extended for effect estimation. Let $H(\psi) \equiv Y - \psi A$, and define $H \equiv Y - \psi_0 A$ so that $H = H(\psi_0)$. Then $E(H \mid A, C, X^*) = E(Y_0 \mid A, C, X^*)$. Our claim is that it is now that H (instead of Y, since H_0 is no longer assumed) behaves like an instrumental variable for the X^* – A association, controlling for C. To see this, first note that by randomization of A within $\{C, X^*\}$, we have $E(H \mid A, C, X^*) = E(Y_0 \mid A, C, X^*) = E(Y_0 \mid C, X^*)$, hence $E(H \mid A, C, X^*) = E(H \mid C, X^*)$. Secondly, by assumption, $H \perp\!\!\!\perp \epsilon^*$, and finally, since X^* is a confounder of the A–Y association, X^* must be correlated with Y_0, and hence H, by definition.

Replacing Y with $H(\psi)$ in the equations given in Theorems 3.2-4 when $q = \dim(\psi)$ (one, under model 3.3) produces a system of estimating equations for ψ, α, and (in the doubly-robust and outcome-regression cases) γ. Unbiasedness of these functions follows analogously to the unbiasedness of the moment equations shown in the proofs for Theorems 3.2-4. Thus, the unknown parameters can be estimated by solving $\mathbb{P}_n U(\psi, \alpha) = 0$ or $\mathbb{P}_n U(\psi, \alpha, \gamma) = 0$. Under mild regularity conditions, the resulting estimator will be consistent and asymptotically normal, and the estimator produced by solving the doubly-robust estimating equations will have these properties provided at least one of $g_A(C; \alpha)$ or $g_Y(C; \gamma)$ is specified correctly.

Optimal choices of functions $\ell(C)$ and $m(C)$ for the robust propensity-score estimators are derived in the appendix. These functions also optimize power when used for hypothesis testing in Section 3.2. We note, however, that these functions depend on parameter estimates from several additional models, and may not necessarily provide efficiency gain depending on the efficiency of the nuisance-parameter estimators. The additional variability introduced by these parameters must be accounted for in finding a suitable variance estimator $\hat{\Omega}_n$ for both testing and estimation. This can be accomplished (as done with γ) by stacking into $U(\psi, \alpha, \gamma)$ the score or estimating equations used to estimate the nuisance parameters that estimates of the functions k, ℓ, and m depend on.

3.8 Extensions to binary and count exposures

The test statistics described in this paper can be extended to binary- and count-exposure cases. Assuming the propensity-score model

$$\text{logit}\,\text{Pr}(A = 1 \mid C, X^*) = g(C; \alpha_1) + \alpha_2^T X^*,$$

for binary A or

$$\log E(A \mid C, X^*) = g(C; \alpha_1) + \alpha_2^T X^*,$$

for count A is correctly specified, we have the following analogous result.

Theorem 3.5. *Let $\ell(C)$ and $m(C)$ each be vector-valued functions of C with linearly-independent elements and dimension $p + q$, where $p = p_1 + p_2$, and let*

$$U(\alpha) \equiv \{\ell(C)Y + m(C)\} \exp(-\alpha_2^T XA)[A - \text{expit}\{\alpha_0^* + g(C; \alpha_1)\}]$$

if A is binary or

$$U(\alpha) \equiv \{\ell(C)Y + m(C)\} [A - \exp\{\alpha_0 + g(C; \alpha_1) + \alpha_2^T X\}]$$

if A is a count. Suppose $U(\alpha)$ is continuously differentiable, $\nabla_\alpha E\{U(\alpha)\} = E\{\nabla_\alpha U(\alpha)\}$, and $\Omega^{-1} E\{\nabla_\alpha U(\alpha)\}$ has full rank, where $\Omega = E\left[U(\alpha)U(\alpha)^T\right]$. Under H_0, the test statistic $\chi^2_{\text{robust}_A} \equiv \min_\alpha n\hat{U}_n(\alpha)^T \hat{\Omega}_n^{-1} \hat{U}_n(\alpha) \xrightarrow{d} \chi^2_q$, for any $\hat{\Omega}_n \xrightarrow{p} \Omega$.

Upon specifying a structural conditional-mean model for the causal effect of A on Y, the moment functions for these tests can be easily adapted to form estimating equations for the average causal effect as was shown for the continuous-exposure case.

3.9 Discussion

We have developed a valid test statistic of the sharp null accounting for the presence of a causal effect when some confounders are contaminated with classical measurement error. This work contributes to the literature on measurement error not only in causal inference, but also in the absence of external information by leveraging causal assumptions to produce a function of the observed data that behaves as an instrumental variable. The tests presented here do not require a causal model to be specified; they only require specification of a conditional mean model of the exposure, outcome, or both. The doubly-robust test only requires one of these models to be correctly specified. The only assumptions required beyond those inherent to the causal inference framework (e.g., no unobserved confounding) are that the conditional mean of exposure is linear in the error-prone confounders and that there are no interaction between the error-prone and error-free confounders.

Sargan tests behave as goodness-of-fit tests, such that when the appropriate models are correctly specified, the tests presented here are powered to detect whether H_0 fits the data. However, the tests are also powered to detect model misspecification, so even in the case where there is no causal effect, the tests will have power to reject. Thus, when our tests reject, it is prudent to supplement them with a Sargan goodness-of-fit test for each model used as described in Section 3.6 in order to ensure the results are not due to poor model fit. Unfortunately, these tests cannot distinguish between nonlinearity in C and nonlinearity in X^* or interactions between X^* and C.

In the multicity application, we tested for an effect of temperature on mortality while accounting for confounding by an error-prone measurement of PM2.5, and discovered evidence of an effect in New York, NY and Stamford, CT. While results from standard tests agreed with our findings in New York, test results in Stamford disagreed with our results, suggesting these standard tests were biased toward the null in this case. In three other cities, the standard tests showed evidence of an effect, while our measurement-error-robust tests did not. This suggested a bias in the standard tests due to measurement error resulting in false positives, and that our method protected us against making such an error.

The work presented here is not without limitations. Though the tests presented here are robust to measurement error of a subset of confounders, at least one true confounder must be measured correctly. While the doubly-robust test allows for misspecification of the conditional mean of the exposure in the error-free confounders, it still requires linearity in the error-prone confounders in the propensity-score model, even when the outcome-regression model is correctly specified. More importantly, while we have managed to avoid the use of parametric models, the assumptions of linearity in the error-contaminated confounders and no interaction in the error-contaminated and error-free confounders on the exposure could be unrealistic in certain settings. In our data application, no goodness-of-fit test rejected at an α level of 0.10 after adding sufficiently many basis functions, however we cannot be certain that the goodness-of-fit tests of the final propensity-score models were powered to detect nonlinearities in the error-contaminated confounders and interactions in the error-contaminated and error-free confounders. Finally, we did not test for lagged effects of temperature, which could have some contribution to the effect of temperature on mortality.

References

AMEMIYA, Y. (1985). Instrumental variable estimator for the nonlinear errors-in-variables model. *Journal of Econometrics* **28** 273–289.

AMEMIYA, Y. (1990). Two-stage instrumental variables estimators for the nonlinear errors-in-variables model. *Journal of Econometrics* **44** 311–332.

AMEMIYA, Y., BROWN, P. and FULLER, W. (1990). Instrumental variable estimation of the nonlinear measurement error model. *Statistical Analysis of Measurement Error Models and Applications* 147–156.

ARMSTRONG, B. (2004). Exposure measurement error: consequences and design issues. *Exposure assessment in occupational and environmental epidemiology* .

ARMSTRONG, B. G. (1990). The effects of measurement errors on relative risk regressions. *American journal of epidemiology* **132** 1176–1184.

AVIN, C., SHPITSER, I. and PEARL, J. (2005). Identifiability of path-specific effects. In *IJCAI-05, Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence*.

BANG, H. and ROBINS, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* **61** 962–973.

BANGSBERG, D. R., HECHT, F. M., CHARLEBOIS, E. D., ZOLOPA, A. R., HOLODNIY, M., SHEINER, L., BAMBERGER, J. D., CHESNEY, M. A. and MOSS, A. (2000). Adherence to protease inhibitors, HIV-1 viral load, and development of drug resistance in an indigent population. *AIDS* **14** 357–366.

BARON, R. M. and KENNY, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology* **51** 1173.

BATESON, T. F., COULL, B. A., HUBBELL, B., ITO, K., JERRETT, M., LUMLEY, T., THOMAS, D., VEDAL, S. and ROSS, M. (2007). Panel discussion review: session three–issues involved in interpretation of epidemiologic analyses–statistical modeling. *Journal of Exposure Science and Environmental Epidemiology* **17** S90–S96.

BATTISTIN, E. and CHESHER, A. (2014). Treatment effect estimation with covariate measurement error. *Journal of Econometrics* **178** 707–715.

BONHOMME, S. and ROBIN, J.-M. (2009). Consistent noisy independent component analysis. *Journal of Econometrics* **149** 12–25.

BUZAS, J. S. and STEFANSKI, L. A. (1996). Instrumental variable estimation in generalized linear measurement error models. *Journal of the American Statistical Association* **91** 999–1006.

CAMPBELL, D. T. (1969). Definitional versus multiple operationalism. *et al* **2** 14–17.

CARROLL, R. and STEFANSKI, L. (1994). Measurement error, instrumental variables and corrections for attenuation with applications to meta-analyses. *Statistics in Medicine* **13** 1265–1282.

CARROLL, R. J., RUPPERT, D., STEFANSKI, L. A. and CRAINICEANU, C. M. (2006). *Measurement error in nonlinear models: a modern perspective*. CRC press.

CHESHER, A. (1991). The effect of measurement error. *Biometrika* **78** 451–462.

CHINA TUBERCULOSIS CONTROL COLLABORATION (1996). Results of directly observed short-course chemotherapy in 112 842 Chinese patients with smear-positive tuberculosis. *The Lancet* **347** 358–362.

COCHRAN, W. G. and RUBIN, D. B. (1973). Controlling bias in observational studies: A review. *Sankhyā: The Indian Journal of Statistics, Series A* 417–446.

COTE, J. A. and BUCKLEY, M. R. (1987). Estimating trait, method, and error variance: Generalizing across 70 construct validation studies. *Journal of Marketing Research* 315–318.

CRAGG, J. G. (1997). Using higher moments to estimate the simple errors-in-variables model. *Rand Journal of Economics* S71–S91.

DAGENAIS, M. G. and DAGENAIS, D. L. (1997). Higher moment estimators for linear regression models with errors in the variables. *Journal of Econometrics* **76** 193–221.

EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics* 1–26.

ELDRED, L. J., WU, A. W., CHAISSON, R. E. and MOORE, R. D. (1998). Adherence to antiretroviral and pneumocystis prophylaxis in HIV disease. *Journal of Acquired Immune Deficiency Syndromes* **18** 117–125.

ERICKSON, T. and WHITED, T. M. (2000). Measurement error and the relationship between investment and q. *Journal of political economy* **108** 1027–1057.

ERICKSON, T. and WHITED, T. M. (2002). Two-step gmm estimation of the errors-in-variables model using high-order moments. *Econometric Theory* **18** 776–799.

ESPELAND, M. A. and HUI, S. L. (1987). A general approach to analyzing epidemiologic data that contain misclassification errors. *Biometrics* 1001–1012.

FAN, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics* 1257–1272.

FAN, J. and TRUONG, Y. K. (1993). Nonparametric regression with errors in variables. *The Annals of Statistics* 1900–1925.

FRISCH, R. (1934). *Statistical confluence analysis by means of complete regression systems*, vol. 5. Universitetets Økonomiske Instituut.

FUJIWARA, P. I., LARKIN, C. and FRIEDEN, T. R. (1997). Directly observed therapy in New York City: History, implementation, results, and challenges. *Clinics in Chest Medicine* **18** 135–148.

FULLER, W. A. (2009). *Measurement error models*, vol. 305. John Wiley & Sons.

FUNG, K. Y. and HOWE, G. R. (1984). Methodological issues in case-control studies iii: The effect of joint misclassification of risk factors and confounding factors upon estimation and power. *International journal of epidemiology* **13** 366–370.

GIFFORD, A., SHIVELY, M., BORMANN, J., TIMBERLAKE, D. and BOZZETTE, S. (1998). Self-reported adherence to combination antiretroviral medication regimens in a community-based sample of HIV-infected adults. In *12th World AIDS Conference*.

GOETGELUK, S., VANSTEELANDT, S. and GOETGHEBEUR, E. (2008). Estimation of controlled direct effects. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70** 1049–1066.

GREENLAND, S. (1980). The effect of misclassification in the presence of covariates. *American Journal of Epidemiology* **112** 564–569.

GRYPARIS, A., PACIOREK, C. J., ZEKA, A., SCHWARTZ, J. and COULL, B. A. (2009). Measurement error caused by spatial misalignment in environmental epidemiology. *Biostatistics* **10** 258–274.

HANSEN, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society* 1029–1054.

HOEK, G., BRUNEKREEF, B., GOLDBOHM, S., FISCHER, P. and VAN DEN BRANDT, P. A. (2002). Association between mortality and indicators of traffic-related air pollution in the netherlands: a cohort study. *The lancet* **360** 1203–1209.

HORVITZ, D. G. and THOMPSON, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* **47** 663–685.

IMAI, K., KEELE, L. and TINGLEY, D. (2010a). A general approach to causal mediation analysis. *Psychological Methods* **15** 309.

IMAI, K., KEELE, L. and YAMAMOTO, T. (2010b). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science* 51–71.

JERRETT, M., BURNETT, R. T., MA, R., POPE III, C. A., KREWSKI, D., NEWBOLD, K. B., THURSTON, G., SHI, Y., FINKELSTEIN, N., CALLE, E. E. ET AL. (2005). Spatial analysis of air pollution and mortality in los angeles. *Epidemiology* **16** 727–736.

KALDOR, J. and CLAYTON, D. (1985). Latent class analysis in chronic disease epidemiology. *Statistics in Medicine* **4** 327–335.

KANG, J. D. and SCHAFER, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science* 523–539.

KAPTEYN, A. and WANSBEEK, T. (1983). Identification in the linear errors in variables model. *Econometrica: Journal of the Econometric Society* 1847–1849.

KELSEY, J. L. (1996). *Methods in observational epidemiology*, vol. 26. Oxford University Press, USA.

KIOUMOURTZOGLOU, M.-A., SPIEGELMAN, D., SZPIRO, A. A., SHEPPARD, L., KAUFMAN, J. D., YANOSKY, J. D., WILLIAMS, R., LADEN, F., HONG, B. and SUH, H. (2014). Exposure measurement error in pm2. 5 health effects studies: A pooled analysis of eight personal exposure validation studies. *Environ Health* **13** 2.

KLEE, V. and MINTY, G. J. (1970). How good is the simplex algorithm. Tech. rep., DTIC Document.

KLEPPER, S. and LEAMER, E. E. (1984). Consistent sets of estimates for regressions with errors in all variables. *Econometrica: Journal of the Econometric Society* 163–183.

LEWBEL, A. (1997). Constructing instruments for regressions with measurement error when no additional data are available, with an application to patents and r&d. *Econometrica: Journal of the Econometric Society* 1201–1213.

LEWBEL, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business & Economic Statistics* **30** 67–80.

MILLS, E. J., NACHEGA, J. B., BUCHAN, I., ORBINSKI, J., ATTARAN, A., SINGH, S., RACHLIS, B., WU, P., COOPER, C., THABANE, L. ET AL. (2006). Adherence to antiretroviral therapy in Sub-Saharan Africa and North America: A meta-analysis. *Journal of the American Medical Association* **296** 679–690.

NEWEY, W. K. and MCFADDEN, D. (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics* **4** 2111–2245.

OGBURN, E. L. and VANDERWEELE, T. J. (2012). On the nondifferential misclassification of a binary confounder. *Epidemiology (Cambridge, Mass.)* **23** 433.

OGBURN, E. L. and VANDERWEELE, T. J. (2013). Bias attenuation results for nondifferentially mismeasured ordinal and coarsened confounders. *Biometrika* **100** 241–248.

O'NEILL, M. S., ZANOBETTI, A. and SCHWARTZ, J. (2003). Modifiers of the temperature and mortality association in seven us cities. *American journal of epidemiology* **157** 1074–1082.

PAL, M. (1980). Consistent moment estimators of regression coefficients in the presence of errors in variables. *Journal of Econometrics* **14** 349–364.

PEARL, J. (2000). *Causality: Models, Reasoning and Inference*. Cambridge University Press, New York. 2nd edition, 2009.

PEARL, J. (2001). Direct and indirect effects. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann Publishers Inc.

PETERSEN, M. L., SINISI, S. E. and VAN DER LAAN, M. J. (2006). Estimation of direct causal effects. *Epidemiology* **17** 276–284.

POP-ELECHES, C., THIRUMURTHY, H., HABYARIMANA, J. P., ZIVIN, J. G., GOLDSTEIN, M. P., DE WALQUE, D., MACKEEN, L., HABERER, J., KIMAIYO, S., SIDLE, J. ET AL. (2011). Mobile phone technologies improve adherence to antiretroviral treatment in a resource-limited setting: A randomized controlled trial of text message reminders. *AIDS (London, England)* **25** 825.

PUETT, R. C., YANOSKY, J. D., HART, J. E., PACIOREK, C. J., SCHWARTZ, J. D., SUH MACINTOSH, H. H., SPEIZER, F. E. and LADEN, F. (2009). Chronic fine and coarse particulate exposure, mortality, and coronary heart disease in the nurses' health study .

RAO, C. R. and ZHAO, L. (1992). Approximation to the distribution of m-estimates in linear models by randomly weighted bootstrap. *Sankhyā: The Indian Journal of Statistics, Series A* 323–331.

RICHARDSON, T. S. (2009). A factorization criterion for acyclic directed mixed graphs. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*. AUAI Press.

RICHARDSON, T. S. and ROBINS, J. M. (2013). Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality. *Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper* .

ROBERTS, K. J. (2000). Barriers to and facilitators of HIV-positive patients' adherence to antiretroviral treatment regimens. *AIDS Patient Care and STDs* **14** 155–168.

ROBINS, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect. *Mathematical Modelling* **7** 1393–1512.

ROBINS, J., SUED, M., LEI-GOMEZ, Q. and ROTNITZKY, A. (2007). Comment: Performance of double-robust estimators when "inverse probability" weights are highly variable. *Statistical Science* 544–559.

ROBINS, J. M. (1989). The analysis of randomized and non-randomized aids treatment trials using a new approach to causal inference in longitudinal studies. *Health service research methodology: a focus on AIDS* **113** 159.

ROBINS, J. M. (1999). Testing and estimation of direct effects by reparameterizing directed acyclic graphs with structural nested models. *Computation, Causation, and Discovery* 349–405.

ROBINS, J. M. (2003). Semantics of causal DAG models and the identification of direct and indirect effects. *Highly Structured Stochastic Systems* 70–81.

ROBINS, J. M. and GREENLAND, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 143–155.

ROBINS, J. M. and RICHARDSON, T. S. (2010). Alternative graphical causal models and the identification of direct effects. *Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures* 103–158.

ROBINS, J. M., RITOV, Y. ET AL. (1997). Toward a curse of dimensionality appropriate (CODA) asymptotic theory for semi-parametric models. *Statistics in Medicine* **16** 285–319.

RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66** 688.

RUBIN, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics* 34–58.

SAMPSON, P. D., SZPIRO, A. A., SHEPPARD, L., LINDSTRÖM, J. and KAUFMAN, J. D. (2011). Pragmatic estimation of a spatio-temporal air quality model with irregular monitoring data. *Atmospheric Environment* **45** 6593–6606.

SARGAN, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica: Journal of the Econometric Society* 393–415.

SCHENNACH, S. M. (2012). Measurement error in nonlinear models: A review. Tech. rep., cemmap working paper, Centre for Microdata Methods and Practice.

SCHENNACH, S. M. (2014). Entropic latent variable integration via simulation. *Econometrica* **82** 345–385.

SCHENNACH, S. M. and HU, Y. (2013). Nonparametric identification and semiparametric estimation of classical measurement error models without side information. *Journal of the American Statistical Association* **108** 177–186.

SCHRIJVER, A. (1998). *Theory of linear and integer programming*. John Wiley & Sons.

Sheppard, L., Burnett, R. T., Szpiro, A. A., Kim, S.-Y., Jerrett, M., Pope III, C. A. and Brunekreef, B. (2012). Confounding and exposure measurement error in air pollution epidemiology. *Air Quality, Atmosphere & Health* **5** 203–216.

Shpitser, I. (2013). Counterfactual graphical models for longitudinal mediation analysis with unobserved confounding. *Cognitive Science* **37** 1011–1035.

Splawa-Neyman, J., Dabrowska, D., Speed, T. et al. (1990). On the application of probability theory to agricultural experiments. essay on principles. section 9. *Statistical Science* **5** 465–472.

Stefanski, L. and Buzas, J. (1995). Instrumental variable estimation in binary regression measurement error models. *Journal of the American Statistical Association* **90** 541–550.

Stuart, A. and Kendall, M. G. (1979). *The advanced theory of statistics*. Macmillan, New York, 4th edition edn.

Suárez, P. G., Watt, C. J., Alarcón, E., Portocarrero, J., Zavala, D., Canales, R., Luelmo, F., Espinal, M. A. and Dye, C. (2001). The dynamics of tuberculosis in response to 10 years of intensive control effort in Peru. *Journal of Infectious Diseases* **184** 473–478.

Szpiro, A. A., Sampson, P. D., Sheppard, L., Lumley, T., Adar, S. D. and Kaufman, J. D. (2010). Predicting intra-urban variation in air pollution concentrations with complex spatio-temporal dependencies. *Environmetrics* **21** 606–631.

Szpiro, A. A., Sheppard, L. and Lumley, T. (2011). Efficient measurement error correction with spatially misaligned data. *Biostatistics* kxq083.

Tang, M. W., Kanki, P. J. and Shafer, R. W. (2012). A review of the virological efficacy of the 4 World Health Organization–recommended tenofovir-containing regimens for initial HIV therapy. *Clinical Infectious Diseases* **54** 862–875.

Taylor, J. M., Wang, Y. and Thiébaut, R. (2005). Counterfactual links to the proportion of treatment effect explained by a surrogate marker. *Biometrics* **61** 1102–1111.

TCHETGEN TCHETGEN, E. J. (2011). On causal mediation analysis with a survival outcome. *The International Journal of Biostatistics* **7** 1–38.

TCHETGEN TCHETGEN, E. J. (2013). Inverse odds ratio-weighted estimation for causal mediation analysis. *Statistics in Medicine* **32** 4567–4580.

TCHETGEN TCHETGEN, E. J. and LIN, S. H. (2012). Robust estimation of pure/natural direct effects with mediator measurement error. *Harvard University Biostatistics Paper Series* **Working Paper 152**.

TCHETGEN TCHETGEN, E. J. and PHIRI, K. (2014). Bounds for pure direct effect. *Epidemiology* **25** 775–776.

TCHETGEN TCHETGEN, E. J. and SHPITSER, I. (2012). Semiparametric theory for causal mediation analysis: Efficiency bounds, multiple robustness and sensitivity analysis. *The Annals of Statistics* **40** 1816–1845.

TCHETGEN TCHETGEN, E. J. and SHPITSER, I. (2014). Semiparametric estimation of models for natural direct and indirect effects. *Biometrika (In press)* .

TCHETGEN TCHETGEN, E. J. and VANDERWEELE, T. J. (2012). On identification of natural direct effects when a confounder of the mediator is directly affected by exposure .

TEN HAVE, T. R., JOFFE, M. M., LYNCH, K. G., BROWN, G. K., MAISTO, S. A. and BECK, A. T. (2007). Causal mediation analyses with rank preserving models. *Biometrics* **63** 926–934.

TIAN, J. and PEARL, J. (2002). A general identification condition for causal effects. In *AAAI/IAAI*.

TZONOU, A., KALDOR, J., SMITH, P., DAY, N. and TRICHOPOULOS, D. (1985). Misclassification in case-control studies with two dichotomous risk factors. *Revue d'épidémiologie et de santé publique* **34** 10–17.

VALERI, L., LIN, X. and VANDERWEELE, T. J. (2014). Mediation analysis when a continuous mediator is measured with error and the outcome follows a generalized linear model. *Statistics in medicine* .

VAN DER LAAN, M. J. and PETERSEN, M. L. (2008). Direct effect models. *The International Journal of Biostatistics* **4** 1–27.

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer.

VANDERWEELE, T. and VANSTEELANDT, S. (2009). Conceptual issues concerning mediation, interventions and composition. *Statistics and its Interface* **2** 457–468.

VANDERWEELE, T. J. (2009). Marginal structural models for the estimation of direct and indirect effects. *Epidemiology* **20** 18–26.

VANDERWEELE, T. J. (2011). Causal mediation analysis with survival data. *Epidemiology (Cambridge, Mass.)* **22** 582.

VANDERWEELE, T. J., VALERI, L. and OGBURN, E. L. (2012). The role of measurement error and misclassification in mediation analysis. *Epidemiology (Cambridge, Mass.)* **23** 561.

VANDERWEELE, T. J. and VANSTEELANDT, S. (2010). Odds ratios for mediation analysis for a dichotomous outcome. *American Journal of Epidemiology* **172** 1339–1348.

VRANCEANU, A. M., SAFREN, S. A., LU, M., COADY, W. M., SKOLNIK, P. R., ROGERS, W. H. and WILSON, I. B. (2008). The relationship of post-traumatic stress disorder and depression to antiretroviral medication adherence in persons with HIV. *AIDS Patient Care and STDs* **22** 313–321.

YANOSKY, J. D., SUH MACINTOSH, H. H. and PACIOREK, C. J. (2008). Predicting chronic fine and coarse particulate exposures using spatiotemporal models for the northeastern and midwestern united states .

ZANOBETTI, A., SCHWARTZ, J. ET AL. (2009). The effect of fine and coarse particulate air pollution on mortality: a national analysis. *Environ Health Perspect* **117** 898–903.

ZEGER, S. L., THOMAS, D., DOMINICI, F., SAMET, J. M., SCHWARTZ, J., DOCKERY, D. and COHEN, A. (2000). Exposure measurement error in time-series studies of air pollution: concepts and consequences. *Environmental health perspectives* **108** 419.

Appendix A

Proofs and theoretical results for Chapter 1

A.1 Identification Result Proof

Proof of Theorem 1.1.

$$\beta_0 \equiv \mathbb{E}[Y(M(\mathbf{C_1}(e'),e),\mathbf{C_1}(e'),e')]$$

$$= \int\limits_{\mathbf{c_0},\mathbf{c_1},m,y} y dF_{Y(M(\mathbf{C_1}(e'),e),\mathbf{C_1}(e'),e'),M(\mathbf{C_1}(e'),e),\mathbf{C_1}(e'),\mathbf{C_0}}(y,m,\mathbf{c_1},\mathbf{c_0})$$

$$= \iint\limits_{\mathbf{c_0},\mathbf{c_1},m,y} y dF_{Y(m,e'),M(\mathbf{c_1},e),\mathbf{C_1}(e')|\mathbf{C_0}}(y,m,\mathbf{c_1}|\mathbf{c_0})dF_{\mathbf{C_0}}(\mathbf{c_0})$$

$$= \iiint\limits_{\mathbf{c_0},\mathbf{c_1},m,y} y dF_{Y(m,e'),\mathbf{C_1}(e')|\mathbf{C_0}}(y,\mathbf{c_1}|\mathbf{c_0})dF_{M(\mathbf{c_1},e)|\mathbf{C_0}}(m|\mathbf{c_0})dF_{\mathbf{C_0}}(\mathbf{c_0}) \tag{A.1}$$

$$= \iiint\limits_{\mathbf{c_0},\mathbf{c_1},m,y} y dF_{Y(m,e'),\mathbf{C_1}(e')|E,\mathbf{C_0}}(y,\mathbf{c_1}|e',\mathbf{c_0})dF_{M(\mathbf{c_1},e)|\mathbf{C_0}}(m|\mathbf{c_0})dF_{\mathbf{C_0}}(\mathbf{c_0}) \tag{A.2}$$

$$= \iiint\limits_{\mathbf{c_0},\mathbf{c_1},m,y} y dF_{Y(m),\mathbf{C_1}|E,\mathbf{C_0}}(y,\mathbf{c_1}|e',\mathbf{c_0})dF_{M(\mathbf{c_1},e)|\mathbf{C_0}}(m|\mathbf{c_0})dF_{\mathbf{C_0}}(\mathbf{c_0}) \tag{A.3}$$

$$= \iiiint\limits_{\mathbf{c_0},\mathbf{c_1},m,y} y dF_{Y(m)|\mathbf{C_1},E,\mathbf{C_0}}(y|\mathbf{c_1},e',\mathbf{c_0})dF_{M(\mathbf{c_1},e)|\mathbf{C_0}}(m|\mathbf{c_0})dF_{\mathbf{C_1}|E,\mathbf{C_0}}(\mathbf{c_1}|e',\mathbf{c_0})dF_{\mathbf{C_0}}(\mathbf{c_0})$$

$$= \iiiint\limits_{\mathbf{c_0},\mathbf{c_1},m,y} y dF_{Y(m)|M,\mathbf{C_1},E,\mathbf{C_0}}(y|m,\mathbf{c_1},e',\mathbf{c_0})dF_{M(\mathbf{c_1},e)|\mathbf{C_1},E,\mathbf{C_0}}(m|\mathbf{c_1},e,\mathbf{c_0})$$

$$\times dF_{\mathbf{C_1}|E,\mathbf{C_0}}(\mathbf{c_1}|e',\mathbf{c_0})dF_{\mathbf{C_0}}(\mathbf{c_0}) \tag{A.4}$$

$$= \iiiint\limits_{\mathbf{c_0},\mathbf{c_1},m,y} y dF_{Y|M,\mathbf{C_1},E,\mathbf{C_0}}(y|m,\mathbf{c_1},e',\mathbf{c_0})dF_{M|\mathbf{C_1},E,\mathbf{C_0}}(m|\mathbf{c_1},e,\mathbf{c_0})$$

$$\times dF_{\mathbf{C_1}|E,\mathbf{C_0}}(\mathbf{c_1}|e',\mathbf{c_0})dF_{\mathbf{C_0}}(\mathbf{c_0}), \tag{A.5}$$

where (A.1) follows from $\{Y(m,e'),\mathbf{C_1}(e')\} \perp\!\!\!\perp M(\mathbf{c_1},e)|\mathbf{C_0}$, (A.2) follows from $\{Y(m,e'),\mathbf{C_1}(e')\} \perp\!\!\!\perp E|\mathbf{C_0}$, (A.3) follows by consistency, (A.4) follows from $Y(m) \perp\!\!\!\perp M|\mathbf{C_1},E,\mathbf{C_0}$ and $M(\mathbf{c_1},e) \perp\!\!\!\perp \{\mathbf{C_1},E\}|\mathbf{C_0}$, and (A.5) follows by consistency. □

A.2 Derivation of Estimation Strategies

A.2.1 Maximum Likelihood Estimator

The maximum likelihood estimator arises from the alternative representation of (1.1):

$$\begin{aligned}&\iiint\limits_{m,\mathbf{c_1},\mathbf{c_0}} \mathbb{E}(Y|m,\mathbf{c_1},e',\mathbf{c_0})dF_{M|\mathbf{C_1},E,\mathbf{C_0}}(m|\mathbf{c_1},e,\mathbf{c_0})dF_{\mathbf{C_1}|E,\mathbf{C_0}}(\mathbf{c_1}|e',\mathbf{c_0})dF_{\mathbf{C_0}}(\mathbf{c_0})\\ &\qquad = \mathbb{E}(\mathbb{E}(\mathbb{E}(\mathbb{E}(Y|M,\mathbf{C_1},e',\mathbf{C_0})|\mathbf{C_1},e,\mathbf{C_0})|e',\mathbf{C_0})).\end{aligned}$$

We replace the inner three expectations with their arguments' means under the empirical laws $\hat{f}_{\mathbf{C_1}|e',\mathbf{C_0}}$, $\hat{f}_{M|\mathbf{C_1},e,\mathbf{C_0}}$, and $\hat{f}_{Y|M,\mathbf{C_1},e',\mathbf{C_1}}$ respectively, and compute the empirical mean. Thus, we have

$$\hat{\beta}_{mle} \equiv \mathbb{P}_n\left\{\hat{\mathbb{E}}(\hat{\mathbb{E}}(\hat{\mathbb{E}}(Y|M,\mathbf{C_1},e',\mathbf{C_0})|\mathbf{C_1},e,\mathbf{C_0})|e',\mathbf{C_0})\right\}.$$

A.2.2 Estimator a

$\hat{\beta}_a$ arises from another alternative representation of (1.1):

$$\begin{aligned}&\iiint\limits_{m,\mathbf{c_1},\mathbf{c_0}} \mathbb{E}(Y|m,\mathbf{c_1},e',\mathbf{c_0})dF_{M|\mathbf{C_1},E,\mathbf{C_0}}(m|\mathbf{c_1},e,\mathbf{c_0})dF_{\mathbf{C_1}|E,\mathbf{C_0}}(\mathbf{c_1}|e',\mathbf{c_0})dF_{\mathbf{C_0}}(\mathbf{c_0})\\ &\qquad = \sum_{e^*\in\{e',e\}} \int\limits_{y,m,\mathbf{c_1},\mathbf{c_0}} y\frac{1_{e'}(e^*)}{f(e'|\mathbf{c_0})}\frac{f(m|\mathbf{c_1},e,\mathbf{c_0})}{f(m|\mathbf{c_1},e^*,\mathbf{c_0})}dF_{Y,M,\mathbf{C_1},E,\mathbf{C_0}}(y,m,\mathbf{c_1},e^*,\mathbf{c_0})\\ &\qquad = \mathbb{E}\left\{Y\frac{1_{e'}(E)}{f(e'|\mathbf{C_0})}\frac{f(M|\mathbf{C_1},e,\mathbf{C_0})}{f(M|\mathbf{C_1},e',\mathbf{C_0})}\right\}.\end{aligned}$$

We simply plug in the empirical laws, $\hat{f}_{E=0|\mathbf{C_0}}$, $\hat{f}_{M|\mathbf{C_1},e,\mathbf{C_0}}$, and $\hat{f}_{M|\mathbf{C_1},e',\mathbf{C_0}}$ for $f_{E=0|\mathbf{C_0}}$, $f_{M|\mathbf{C_1},e,\mathbf{C_0}}$, and $f_{M|\mathbf{C_1},e',\mathbf{C_0}}$ respectively, and compute the empirical mean. Thus, we have

$$\hat{\beta}_a \equiv \mathbb{P}_n\left\{Y\frac{1_{e'}(E)}{\hat{f}(e'|\mathbf{C_0})}\frac{\hat{f}(M|\mathbf{C_1},e,\mathbf{C_0})}{\hat{f}(M|\mathbf{C_1},e',\mathbf{C_0})}\right\}.$$

A.2.3 Estimator b

$\hat{\beta}_b$ arises from a third representation of (1.1):

$$\begin{aligned}\iiint\limits_{m,\mathbf{c_1},\mathbf{c_0}} &\mathbb{E}(Y|m,\mathbf{c_1},e',\mathbf{c_0})dF_{M|\mathbf{C_1},E,\mathbf{C_0}}(m|\mathbf{c_1},e,\mathbf{c_0})dF_{\mathbf{C_1}|E,\mathbf{C_0}}(\mathbf{c_1}|e',\mathbf{c_0})dF_{\mathbf{C_0}}(\mathbf{c_0}) \\ &= \sum_{e^*\in\{e',e\}} \int\limits_{m,\mathbf{c_1},\mathbf{c_0}} \mathbb{E}(Y|M,\mathbf{C_1},e',\mathbf{C_0})\frac{1_e(e^*)}{f(e^*|\mathbf{c_0})}\frac{f(\mathbf{c_1}|e',\mathbf{c_0})}{f(\mathbf{c_1}|e^*,\mathbf{c_0})}dF_{M,\mathbf{C_1},E,\mathbf{C_0}}(m,\mathbf{c_1},e^*,\mathbf{c_0}) \\ &= \mathbb{E}\left[\frac{1_e(E)}{f(e|\mathbf{C_0})}\frac{f(\mathbf{C_1}|e',\mathbf{C_0})}{f(\mathbf{C_1}|e,\mathbf{C_0})}\mathbb{E}(Y|M,\mathbf{C_1},e',\mathbf{C_0})\right].\end{aligned}$$

Again, we plug in the empirical laws $\hat{f}_{\mathbf{C_1}|e',\mathbf{C_0}}$, $\hat{f}_{\mathbf{C_1}|e,\mathbf{C_0}}$, and $\hat{f}_{E=1|\mathbf{C_0}}$ for $f_{\mathbf{C_1}|e',\mathbf{C_0}}$, $f_{\mathbf{C_1}|e,\mathbf{C_0}}$, and $f_{E=1|\mathbf{C_0}}$, respectively, replace $\mathbb{E}(Y|M,\mathbf{C_1},e',\mathbf{C_0})$ with $\hat{\mathbb{E}}(Y|M,\mathbf{C_1},e',\mathbf{C_0})$, the expectation of Y under the empirical law $\hat{f}_{Y|M,\mathbf{C_1},e',\mathbf{C_0}}$, and compute the empirical mean. Thus, we have

$$\hat{\beta}_b \equiv \mathbb{P}_n\left\{\frac{1_e(E)}{\hat{f}(e|\mathbf{C_0})}\frac{\hat{f}(\mathbf{C_1}|e',\mathbf{C_0})}{\hat{f}(\mathbf{C_1}|e,\mathbf{C_0})}\hat{\mathbb{E}}(Y|M,\mathbf{C_1},e',\mathbf{C_0})\right\}.$$

We develop the multiply-robust estimator and prove its robustness properties in the following two sections.

A.3 Derivation of the Influence Function

Theorem A.1. *The efficient influence function of β_0 in model $\mathcal{M}_{nonpar}$ is given by*

$$\begin{aligned}V^{eff}(\beta_0) =& \frac{1_{e'}(E)f(M|e,\mathbf{C_1},\mathbf{C_0})}{f(M|e',\mathbf{C_1},\mathbf{C_0})f(e'|\mathbf{C_0})}\{Y - B(M,\mathbf{C_1},e',\mathbf{C_0})\} \\ &+ \frac{1_e(E)f(\mathbf{C_1}|e',\mathbf{C_0})}{f(\mathbf{C_1}|e,\mathbf{C_0})f(e|\mathbf{C_0})}\{B(M,\mathbf{C_1},e',\mathbf{C_0}) - B'(\mathbf{C_1},e',e,\mathbf{C_0})\} \\ &+ \frac{1_{e'}(E)}{f(e'|\mathbf{C_0})}\{B'(\mathbf{C_1},e',e,\mathbf{C_0}) - B''(e',e,\mathbf{C_0})\} + \{B''(e',e,\mathbf{C_0}) - \beta_0\},\end{aligned}$$

implying that the asymptotic variance of a regular, asymptotically linear (RAL) estimator of β_0 in model $\mathcal{M}_{nonpar}$ can be no smaller than $\mathbb{E}\{V^{eff}(\beta_0)^2\}^{-1}$, the semiparametric efficiency bound for the model.

$\hat{\beta}_{mr}$ is obtained simply by solving the estimating equation $V^{eff}(\beta_0)$ for β_0. Since our model is nonparametric, the asymptotic variance is the same for any estimator in $\mathcal{M}_{nonpar}$ so long as it is RAL. Furthermore, since all such estimators share the common influence function $V^{eff}(\beta_0)$, they also share a common asymptotic expansion, viz. $n^{1/2}(\hat{\beta}_0 - \beta_0) = n^{1/2}\mathbb{P}_n V^{eff}(\beta_0) + o_p(1)$, where $\mathbb{P}_n$ denotes the empirical mean.

Proof. Let ν denote the appropriate dominating measure or product measure corresponding to each combination of random variables. Let $F_{\mathbf{O};t} = F_{Y|M,C,E,\mathbf{C_0};t}F_{M|\mathbf{C_1},E,\mathbf{C_0};t}F_{\mathbf{C_1}|E,\mathbf{C_0};t}F_{E|\mathbf{C_0};t}F_{\mathbf{C_0};t}$ denote a one-dimensional regular parametric submodel of $\mathcal{M}_{nonpar}$ with $F_{\mathbf{O},0} = F_{\mathbf{O}}$, and let

$$\begin{aligned}\beta_t &= \beta_0(F_{\mathbf{O};t}) = \mathbb{E}_t(Y(M(e, \mathbf{C_1}(e')), \mathbf{C_1}(e'), e')) \\ &= \int\limits_{m,\mathbf{c_1},\mathbf{c_0}} \mathbb{E}_t(Y|m, \mathbf{c_1}, e', \mathbf{c_0}) f_t(M = m|\mathbf{c_1}, e, \mathbf{c_0}) f_t(\mathbf{C_1} = \mathbf{c_1}|e', \mathbf{c_0}) f_t(\mathbf{C_0} = \mathbf{c_0}) d\nu(m, \mathbf{c_1}, \mathbf{c_0})\end{aligned}$$

and $U_{\mathbf{O}} = \frac{\nabla_{t=0} f_t(\mathbf{O})}{f(\mathbf{O})}$ be the score for $\mathbf{O}$. Then

$$\left.\frac{\partial \beta_t}{\partial t}\right|_{t=0} =$$

$$\int\limits_{m,\mathbf{c_1},\mathbf{c_0}} \nabla_{t=0}\mathbb{E}_t(Y|m, \mathbf{c_1}, e', \mathbf{c_0}) f(M = m|\mathbf{c_1}, e, \mathbf{c_0}) f(\mathbf{C_1} = \mathbf{c_1}|e', \mathbf{c_0}) f(\mathbf{C_0} = \mathbf{c_0}) d\nu(m, \mathbf{c_1}, \mathbf{c_0}) \tag{A.6}$$

$$+ \int\limits_{m,\mathbf{c_1},\mathbf{c_0}} \mathbb{E}(Y|m, \mathbf{c_1}, e', \mathbf{c_0}) \nabla_{t=0} f_t(M = m|\mathbf{c_1}, e, \mathbf{c_0}) f(\mathbf{C_1} = \mathbf{c_1}|e', \mathbf{c_0}) f(\mathbf{C_0} = \mathbf{c_0}) d\nu(m, \mathbf{c_1}, \mathbf{c_0}) \tag{A.7}$$

$$+ \int\limits_{m,\mathbf{c_1},\mathbf{c_0}} \mathbb{E}(Y|m, \mathbf{c_1}, e', \mathbf{c_0}) f(M = m|\mathbf{c_1}, e, \mathbf{c_0}) \nabla_{t=0} f_t(\mathbf{C_1} = \mathbf{c_1}|e', \mathbf{c_0}) f(\mathbf{C_0} = \mathbf{c_0}) d\nu(m, \mathbf{c_1}, \mathbf{c_0}) \tag{A.8}$$

$$+ \int\limits_{m,\mathbf{c_1},\mathbf{c_0}} \mathbb{E}(Y|m, \mathbf{c_1}, e', \mathbf{c_0}) f(M = m|\mathbf{c_1}, e, \mathbf{c_0}) f(\mathbf{C_1} = \mathbf{c_1}|e', \mathbf{c_0}) \nabla_{t=0} f_t(\mathbf{C_0} = \mathbf{c_0}) d\nu(m, \mathbf{c_1}, \mathbf{c_0}), \tag{A.9}$$

where

$$\begin{aligned}(A.6) = \int\limits_{m,\mathbf{c_1},\mathbf{c_0}} &\nabla_{t=0}\mathbb{E}_t(Y|m, \mathbf{c_1}, e', \mathbf{c_0}) f(M = m|\mathbf{c_1}, e, \mathbf{c_0}) f(\mathbf{C_1} = \mathbf{c_1}|e', \mathbf{c_0}) f(\mathbf{C_0} = \mathbf{c_0}) \\ &\times d\nu(m, \mathbf{c_1}, \mathbf{c_0})\end{aligned}$$

$$
\begin{aligned}
&= \int\limits_{m,\mathbf{c_1},\mathbf{c_0}} \int_y y \left\{ \frac{\nabla_{t=0} f_t(y,m,\mathbf{c_1},e',\mathbf{c_0})}{f(m,\mathbf{c_1},e',\mathbf{c_0})} - \frac{f(y,m,\mathbf{c_1},e',\mathbf{c_0})\nabla_{t=0} f_t(m,\mathbf{c_1},e',\mathbf{c_0})}{f(m,\mathbf{c_1},e',\mathbf{c_0})^2} \right\} d\nu(y) \\
&\quad \times f(M=m|\mathbf{c_1},e,\mathbf{c_0}) f(\mathbf{C_1}=\mathbf{c_1}|e',\mathbf{c_0}) f(\mathbf{C_0}=\mathbf{c_0}) d\nu(m,\mathbf{c_1},\mathbf{c_0}) \\
&= \int\limits_{y,m,\mathbf{c_1},\mathbf{c_0}} \left\{ y \frac{\nabla_{t=0} f_t(y,m,\mathbf{c_1},e',\mathbf{c_0})}{f(m,\mathbf{c_1},e',\mathbf{c_0})} - \frac{\nabla_{t=0} f_t(m,\mathbf{c_1},e',\mathbf{c_0})}{f(m,\mathbf{c_1},e',\mathbf{c_0})} \mathbb{E}(Y|m,\mathbf{c_1},e',\mathbf{c_0}) \right. \\
&\quad \left. \times f(y|m,\mathbf{c_1},e',\mathbf{c_0}) \right\} f(M=m|\mathbf{c_1},e,\mathbf{c_0}) f(\mathbf{C_1}=\mathbf{c_1}|e',\mathbf{c_0}) f(\mathbf{C_0}=\mathbf{c_0}) d\nu(y,m,\mathbf{c_1},\mathbf{c_0}) \\
&= \int\limits_{y,m,\mathbf{c_1},e^*,\mathbf{c_0}} \left\{ y \frac{\nabla_{t=0} f_t(y,m,\mathbf{c_1},e^*,\mathbf{c_0})}{f(m,\mathbf{c_1},e',\mathbf{c_0})} - \frac{\nabla_{t=0} f_t(m,\mathbf{c_1},e^*,\mathbf{c_0})}{f(m,\mathbf{c_1},e',\mathbf{c_0})} \mathbb{E}(Y|m,\mathbf{c_1},e',\mathbf{c_0}) \right. \\
&\quad \left. \times f(y|m,\mathbf{c_1},e',\mathbf{c_0}) \right\} 1_{e'}(e^*) f(M=m|\mathbf{c_1},e,\mathbf{c_0}) f(\mathbf{C_1}=\mathbf{c_1}|e',\mathbf{c_0}) f(\mathbf{C_0}=\mathbf{c_0}) \\
&\quad \times d\nu(y,m,\mathbf{c_1},e^*,\mathbf{c_0}) \\
&= \mathbb{E}\left[\frac{1_{e'}(E) f(M|\mathbf{C_1},e',\mathbf{C_0}) f(\mathbf{C_1}|e',\mathbf{C_0}) f(\mathbf{C_0})}{f(Y,M,\mathbf{C_1},E,\mathbf{C_0})} \times \left\{ Y \frac{\nabla_{t=0} f_t(Y,M,\mathbf{C_1},E,\mathbf{C_0})}{f(M,\mathbf{C_1},e',\mathbf{C_0})} \right.\right. \\
&\quad \left.\left. - f(Y|M,\mathbf{C_1},e',\mathbf{C_0}) \frac{\nabla_{t=0} f_t(M,\mathbf{C_1},E,\mathbf{C_0})}{f(M,\mathbf{C_1},e',\mathbf{C_0})} B(M,\mathbf{C_1},e',\mathbf{C_0}) \right\} \right] \\
&= \mathbb{E}\left[\frac{1_{e'}(E) f(M|\mathbf{C_1},e,\mathbf{C_0}) f(\mathbf{C_1}|e',\mathbf{C_0}) f(\mathbf{C_0})}{f(Y,M,\mathbf{C_1},E,\mathbf{C_0}) f(M,\mathbf{C_1},e',\mathbf{C_0})} \{ Y \nabla_{t=0} f_t(Y,M,\mathbf{C_1},E,\mathbf{C_0}) \right. \\
&\quad - [\nabla_{t=0} f_t(Y,M,\mathbf{C_1},E,\mathbf{C_0}) - f(M,\mathbf{C_1},e',\mathbf{C_0}) \nabla_{t=0} f_t(Y|M,\mathbf{C_1},E,\mathbf{C_0})] \\
&\quad \left. \times B(M,\mathbf{C_1},e',\mathbf{C_0}) \} \right] \\
&= \mathbb{E}\left[\frac{\nabla_{t=0} f_t(Y,M,\mathbf{C_1},E,\mathbf{C_0})}{f(Y,M,\mathbf{C_1},E,\mathbf{C_0})} \frac{1_{e'}(E) f(M|\mathbf{C_1},e,\mathbf{C_0})}{f(M|\mathbf{C_1},e',\mathbf{C_0}) f(E=e'|\mathbf{C_0})} \{ Y - B(M,\mathbf{C_1},e',\mathbf{C_0}) \} \right] \\
&\quad + \int\limits_{m,\mathbf{c_1},\mathbf{c_0}} f(\mathbf{c_1}|e',\mathbf{c_0}) f(m|\mathbf{c_1},e,\mathbf{c_0}) f(\mathbf{c_0}) \nabla_{t=0} \left\{ \int_y f_t(y|m,\mathbf{c_1},e',\mathbf{c_0}) d\nu(y) \right\} \\
&\quad \times B(m,\mathbf{c_1},e',\mathbf{c_0}) d\nu(m,\mathbf{c_1},\mathbf{c_0}) \\
&= \mathbb{E}\left[U_{\mathbf{O}} \frac{1_{e'}(E) f(M|\mathbf{C_1},e,\mathbf{C_0})}{f(M|\mathbf{C_1},e',\mathbf{C_0}) f(E=e'|\mathbf{C_0})} \{ Y - B(M,\mathbf{C_1},e',\mathbf{C_0}) \} \right],
\end{aligned}
$$

$$
\begin{aligned}
(A.7) &= \int\limits_{m,\mathbf{c_1},\mathbf{c_0}} \mathbb{E}(Y|m,\mathbf{c_1},e',\mathbf{c_0}) \left\{ \frac{\nabla_{t=0} f_t(m,\mathbf{c_1},e,\mathbf{c_0})}{f(\mathbf{c_1},e,\mathbf{c_0})} - \frac{\nabla_{t=0} f_t(\mathbf{c_1},e,\mathbf{c_0}) f(m,\mathbf{c_1},e,\mathbf{c_0})}{f(\mathbf{c_1},e,\mathbf{c_0})^2} \right\} \\
&\quad \times f(\mathbf{c_1}|e',\mathbf{c_0}) f(\mathbf{c_0}) d\nu(m,\mathbf{c_1},\mathbf{c_0}) \\
&= \int\limits_{m,\mathbf{c_1},\mathbf{c_0}} \mathbb{E}(Y|m,\mathbf{c_1},e',\mathbf{c_0}) \frac{\nabla_{t=0} f_t(m,\mathbf{c_1},e,\mathbf{c_0})}{f(\mathbf{c_1},e,\mathbf{c_0})} f(\mathbf{c_1}|e',\mathbf{c_0}) f(\mathbf{c_0}) d\nu(m,\mathbf{c_1},\mathbf{c_0}) \\
&\quad - \int\limits_{\mathbf{c_1},\mathbf{c_0}} \frac{\nabla_{t=0} f_t(\mathbf{c_1},e,\mathbf{c_0})}{f(\mathbf{c_1},e,\mathbf{c_0})} \mathbb{E}(\mathbb{E}(Y|m,\mathbf{c_1},e',\mathbf{c_0})|\mathbf{c_1},e,\mathbf{c_0}) f(\mathbf{c_1}|e',\mathbf{c_0}) f(\mathbf{c_0}) d\nu(\mathbf{c_1},\mathbf{c_0})
\end{aligned}
$$

$$
\begin{aligned}
&= \int\limits_{m,\mathbf{c_1},\mathbf{c_0}} \frac{f(\mathbf{c_1}|e',\mathbf{c_0})f(\mathbf{c_0})}{f(\mathbf{c_1},e,\mathbf{c_0})}\big\{\nabla_{t=0}f_t(m,\mathbf{c_1},e,\mathbf{c_0})\mathbb{E}(Y|m,\mathbf{c_1},e',\mathbf{c_0}) \\
&\quad - \nabla_{t=0}f_t(\mathbf{c_1},e,\mathbf{c_0})f(m|\mathbf{c_1},e,\mathbf{c_0})B'(\mathbf{c_1},e',e,\mathbf{c_0})\big\}d\nu(m,\mathbf{c_1},\mathbf{c_0}) \\
&= \int\limits_{m,\mathbf{c_1},\mathbf{c_0}} \frac{f(\mathbf{c_1}|e',\mathbf{c_0})}{f(\mathbf{c_1}|e,\mathbf{c_0})f(e|\mathbf{c_0})}\big\{\nabla_{t=0}f_t(m,\mathbf{c_1},e,\mathbf{c_0})\mathbb{E}(Y|m,\mathbf{c_1},e',\mathbf{c_0}) \\
&\quad - \left[\nabla_{t=0}f_t(m,\mathbf{c_1},e,\mathbf{c_0}) - f(\mathbf{c_1},e,\mathbf{c_0})\nabla_{t=0}f_t(m|\mathbf{c_1},e,\mathbf{c_0})\right]B'(\mathbf{c_1},e',e,\mathbf{c_0})\big\}d\nu(m,\mathbf{c_1},\mathbf{c_0}) \\
&= \int\limits_{m,\mathbf{c_1},\mathbf{c_0}} \nabla_{t=0}f_t(m,\mathbf{c_1},e,\mathbf{c_0})\frac{f(\mathbf{c_1}|e',\mathbf{c_0})}{f(\mathbf{c_1}|e,\mathbf{c_0})f(e|\mathbf{c_0})}\left\{E(Y|m,\mathbf{c_1},e',\mathbf{c_0}) - B'(\mathbf{c_1},e',e,\mathbf{c_0})\right\} \\
&\quad \times d\nu(m,\mathbf{c_1},\mathbf{c_0}) + \int\limits_{\mathbf{c_1},\mathbf{c_0}} f(\mathbf{c_1}|e',\mathbf{c_0})f(\mathbf{c_0})\nabla_{t=0}\int_m f_t(m|\mathbf{c_1},e,\mathbf{c_0})d\nu(m)B'(\mathbf{c_1},e',e,\mathbf{c_0}) \\
&\quad \times d\nu(\mathbf{c_1},\mathbf{c_0}) \\
&= \int\limits_{m,\mathbf{c_1},\mathbf{c_0}} \Bigg\{\int_y f(y|m,\mathbf{c_1},e,\mathbf{c_0})d\nu(y)\nabla_{t=0}f_t(m,\mathbf{c_1},e,\mathbf{c_0}) \\
&\quad +\nabla_{t=0}\int_y f_t(y|m,\mathbf{c_1},e,\mathbf{c_0})d\nu(y)f(m,\mathbf{c_1},e,\mathbf{c_0})\Bigg\}\frac{f(\mathbf{c_1}|e',\mathbf{c_0})}{f(\mathbf{c_1}|e,\mathbf{c_0})f(e|\mathbf{c_0})} \\
&\quad \times \left\{\mathbb{E}(Y|m,\mathbf{c_1},e',\mathbf{c_0}) - B'(\mathbf{c_1},e',e,\mathbf{c_0})\right\}d\nu(m,\mathbf{c_1},\mathbf{c_0}) \\
&= \int\limits_{y,m,\mathbf{c_1},\mathbf{c_0}} \nabla_{t=0}f_t(y,m,\mathbf{c_1},e,\mathbf{c_0})\frac{f(\mathbf{c_1}|e',\mathbf{c_0})}{f(\mathbf{c_1}|e,\mathbf{c_0})f(e|\mathbf{c_0})} \\
&\quad \times \left\{\mathbb{E}(Y|m,\mathbf{c_1},e',\mathbf{c_0}) - B'(\mathbf{c_1},e',e,\mathbf{c_0})\right\}d\nu(y,m,\mathbf{c_1},\mathbf{c_0}) \\
&= \int\limits_{y,m,\mathbf{c_1},e^*,\mathbf{c_0}} \nabla_{t=0}f_t(y,m,\mathbf{c_1},e^*,\mathbf{c_0})\frac{1_e(e^*)f(\mathbf{c_1}|e',\mathbf{c_0})}{f(\mathbf{c_1}|e,\mathbf{c_0})f(e|\mathbf{c_0})} \\
&\quad \times \left\{\mathbb{E}(Y|m,\mathbf{c_1},e',\mathbf{c_0}) - B'(\mathbf{c_1},e',e,\mathbf{c_0})\right\}d\nu(y,m,\mathbf{c_1},e^*,\mathbf{c_0}) \\
&=\mathbb{E}\left[U_{\mathbf{O}}\frac{1_e(E)f(\mathbf{C_1}|e',\mathbf{C_0})}{f(\mathbf{C_1}|e,\mathbf{C_0})f(e|\mathbf{C_0})}\left\{\mathbb{E}(Y|M,\mathbf{C_1},e',\mathbf{C_0}) - B'(\mathbf{C_1},e',e,\mathbf{C_0})\right\}\right],
\end{aligned}
$$

$$
\begin{aligned}
(A.8) &= \int\limits_{m,\mathbf{c_1},\mathbf{c_0}} \mathbb{E}(Y|m,\mathbf{c_1},e',\mathbf{c_0})f(m|\mathbf{c_1},e,\mathbf{c_0})\left\{\frac{\nabla_{t=0}f_t(\mathbf{c_1},e',\mathbf{c_0})}{f(e',\mathbf{c_0})}\right. \\
&\quad \left.-\frac{\nabla_{t=0}f_t(e',\mathbf{c_0})f(\mathbf{c_1},e',\mathbf{c_0})}{f(e',\mathbf{c_0})^2}\right\}f(\mathbf{c_0})d\nu(m,\mathbf{c_1},\mathbf{c_0}) \\
&= \int\limits_{\mathbf{c_1},\mathbf{c_0}} \mathbb{E}(\mathbb{E}(Y|M,\mathbf{C_1},e',\mathbf{C_0})|\mathbf{c_1},e,\mathbf{c_0})\left\{\frac{\nabla_{t=0}f_t(\mathbf{c_1},e',\mathbf{c_0})}{f(e',\mathbf{c_0})} - \frac{\nabla_{t=0}f_t(e',\mathbf{c_0})}{f(e',\mathbf{c_0})}f(\mathbf{c_1}|e',\mathbf{c_0})\right\} \\
&\quad \times f(\mathbf{c_0})d\nu(\mathbf{c_1},\mathbf{c_0}) \\
&= \int\limits_{\mathbf{c_1},\mathbf{c_0}} B'(\mathbf{c_1},e',e,\mathbf{c_0})\frac{\nabla_{t=0}f_t(\mathbf{c_1},e',\mathbf{c_0})}{f(e',\mathbf{c_0})}f(\mathbf{c_0})d\nu(\mathbf{c_1},\mathbf{c_0})
\end{aligned}
$$

$$
\begin{aligned}
&-\int_{\mathbf{c_0}} \mathbb{E}(\mathbb{E}(\mathbb{E}(Y|M,\mathbf{C_1},e',\mathbf{C_0})|\mathbf{C_1},e,\mathbf{C_0})|e',\mathbf{c_0})\frac{\nabla_{t=0}f_t(e',\mathbf{c_0})}{f(e',\mathbf{c_0})}f(\mathbf{c_0})d\nu(\mathbf{c_0})\\
=&\int\limits_{\mathbf{c_1},\mathbf{c_0}} \frac{f(\mathbf{c_0})}{f(e',\mathbf{c_0})}\left\{\nabla_{t=0}f_t(\mathbf{c_1},e',\mathbf{c_0})B'(\mathbf{c_1},e',e,\mathbf{c_0})\right.\\
&\left.-\nabla_{t=0}f_t(e',\mathbf{c_0})f(\mathbf{c_1}|e',\mathbf{c_0})B''(e',e,\mathbf{c_0})\right\}d\nu(\mathbf{c_1},\mathbf{c_0})\\
=&\int\limits_{\mathbf{c_1},\mathbf{c_0}} \frac{1}{f(e'|\mathbf{c_0})}\left\{\nabla_{t=0}f_t(\mathbf{c_1},e',\mathbf{c_0})B'(\mathbf{c_1},e',e,\mathbf{c_0})\right.\\
&\left.-\left[\nabla_{t=0}f_t(\mathbf{c_1},e',\mathbf{c_0})-\nabla_{t=0}f_t(\mathbf{c_1}|e',\mathbf{c_0})f(e',\mathbf{c_0})\right]B''(e',e,\mathbf{c_0})\right\}d\nu(\mathbf{c_1},\mathbf{c_0})\\
=&\int\limits_{\mathbf{c_1},\mathbf{c_0}} \frac{1}{f(e'|\mathbf{c_0})}\nabla_{t=0}f_t(\mathbf{c_1},e',\mathbf{c_0})\left\{B'(\mathbf{c_1},e',e,\mathbf{c_0})-B''(e',e,\mathbf{c_0})\right\}d\nu(\mathbf{c_1},\mathbf{c_0})\\
&+\int_{\mathbf{c_0}} f(\mathbf{c_0})\nabla_{t=0}\int_{\mathbf{c_1}} f_t(\mathbf{c_1}|e',\mathbf{c_0})d\nu(\mathbf{c_1})B''(e',e,\mathbf{c_0})d\nu(\mathbf{c_0})\\
=&\int\limits_{\mathbf{c_1},\mathbf{c_0}} \frac{1}{f(e'|\mathbf{c_0})}\left\{\int\limits_{y,m} f(y,m|\mathbf{c_1},e',\mathbf{c_0})d\nu(y,m)\nabla_{t=0}f_t(\mathbf{c_1},e',\mathbf{c_0})\right.\\
&\left.+\nabla_{t=0}\int\limits_{y,m} f_t(y,m|\mathbf{c_1},e',\mathbf{c_0})d\nu(y,m)f(\mathbf{c_1},e',\mathbf{c_0})\right\}\left\{B'(\mathbf{c_1},e',e,\mathbf{c_0})-B''(e',e,\mathbf{c_0})\right\}\\
&\times d\nu(\mathbf{c_1},\mathbf{c_0})\\
=&\int\limits_{y,m,\mathbf{c_1},\mathbf{c_0}} \frac{\nabla_{t=0}f_t(y,m,\mathbf{c_1},e',\mathbf{c_0})}{f(e'|\mathbf{c_0})}\left\{B'(\mathbf{c_1},e',e,\mathbf{c_0})-B''(e',e,\mathbf{c_0})\right\}d\nu(y,m,\mathbf{c_1},\mathbf{c_0})\\
=&\int\limits_{y,m,\mathbf{c_1},e^*,\mathbf{c_0}} \nabla_{t=0}f_t(y,m,\mathbf{c_1},e^*,\mathbf{c_0})\frac{1_{e'}(e^*)}{f(e'|\mathbf{c_0})}\left\{B'(\mathbf{c_1},e',e,\mathbf{c_0})-B''(e',e,\mathbf{c_0})\right\}\\
&\times d\nu(y,m,\mathbf{c_1},e^*,\mathbf{c_0})\\
=&\mathbb{E}\left[U_{\mathbf{O}}\frac{1_e(E')}{f(e'|\mathbf{C_0})}\left\{B'(\mathbf{c_1},e',e,\mathbf{C_0})-B''(e',e,\mathbf{C_0})\right\}\right],
\end{aligned}
$$

and

$$
\begin{aligned}
(A.9)=&\int_{\mathbf{c_0}} \mathbb{E}(\mathbb{E}(\mathbb{E}(Y|M,\mathbf{C_1},e',\mathbf{C_0})|\mathbf{C_1},e,\mathbf{C_0})|e',\mathbf{C_0})\nabla_{t=0}f_t(\mathbf{c_0})d\nu(\mathbf{c_0})-\beta_0\mathbb{E}U_{\mathbf{O}}\\
=&\int_{\mathbf{c_0}}\left\{\int\limits_{y,m,\mathbf{c_1},e^*} f(y,m,\mathbf{c_1},e^*|\mathbf{c_0})d\nu(y,m,\mathbf{c_1},e^*)\nabla_{t=0}f_t(\mathbf{c_0})\right.
\end{aligned}
$$

$$
\begin{aligned}
&+\nabla_{t=0} \int\limits_{y,m,\mathbf{c_1},e^*} f_t(y,m,\mathbf{c_1},e^*|\mathbf{c_0})d\nu(y,m,\mathbf{c_1},e^*)f(\mathbf{c_0}) \Bigg\} B''(e',e,\mathbf{c_0})d\nu(\mathbf{c_0}) - \mathbb{E}[U_{\mathbf{O}}\beta_0] \\
= &\int\limits_{y,m,\mathbf{c_1},e^*,\mathbf{c_0}} \nabla_{t=0} f_t(y,m,\mathbf{c_1},e^*,\mathbf{c_0})B''(e',e,\mathbf{c_0})d\nu(y,m,\mathbf{c_1},e^*,\mathbf{c_0}) - \mathbb{E}[U_{\mathbf{O}}\beta_0] \\
= &\mathbb{E}\left[U_{\mathbf{O}}\left\{B''(e',e,\mathbf{C_0}) - \beta_0\right\}\right].
\end{aligned}
$$

Thus, $\frac{\partial \beta_t}{\partial t}\Big|_{t=0} = \mathbb{E}[U_{\mathbf{O}}V^{eff}(\beta_0)]$ where

$$
\begin{aligned}
V^{eff}(\beta_0) = &\frac{1_{e'}(E)f(M|e,\mathbf{C_1},\mathbf{C_0})}{f(M|e',\mathbf{C_1},\mathbf{C_0})f(e'|\mathbf{C_0})}\left\{Y - B(M,\mathbf{C_1},e',\mathbf{C_0})\right\} \\
&+ \frac{1_e(E)f(\mathbf{C_1}|e',\mathbf{C_0})}{f(\mathbf{C_1}|e,\mathbf{C_0})f(e|\mathbf{C_0})}\left\{B(M,\mathbf{C_1},e',\mathbf{C_0}) - B'(\mathbf{C_1},e',e,\mathbf{C_0})\right\} \\
&+ \frac{1_{e'}(E)}{f(e'|\mathbf{C_0})}\left\{B'(\mathbf{C_1},e',e,\mathbf{C_0}) - B''(e',e,\mathbf{C_0})\right\} + \left\{B''(e',e,\mathbf{C_0}) - \beta_0\right\},
\end{aligned}
$$

so if a RAL estimator exists, then $V^{eff}(\beta_0)$ is the corresponding influence function. It is efficient because the model $\mathcal{M}_{nonpar}$ is nonparametric. □

A.4 Multiple-Robustness of the Efficient Influence Function

Let $\tilde{B}$, $\tilde{\boldsymbol{\theta}}_{\boldsymbol{M}} = \{\tilde{M}^{ratio}, \tilde{\mathbb{E}}[\tilde{B}(M,\mathbf{C_1},e',\mathbf{C_0})|\mathbf{C_1},e,\mathbf{C_0}]\}$, $\tilde{\boldsymbol{\theta}}_{\mathbf{C_1}} = \{\tilde{C}_1^{ratio}, \tilde{\mathbb{E}}[\tilde{B}'(\mathbf{C_1},e,\mathbf{c_0})|e',\mathbf{C_0}]\}$, and $\tilde{f}_{E|\mathbf{C_0}}$ denote limits of the estimators using the working models B^W, $\boldsymbol{\theta}_{\boldsymbol{M}}^{\boldsymbol{W}}$, $\boldsymbol{\theta}_{\mathbf{C_1}}^{\boldsymbol{W}}$, and $f_{E|\mathbf{C_0}}^W$. We have established the following multiply-robust property of V^{eff}:

Theorem A.2. *The estimating equation* $V^{eff}(\beta_0, \tilde{B}, \tilde{\boldsymbol{\theta}}_{\mathbf{M}}, \tilde{\boldsymbol{\theta}}_{\mathbf{C_1}}, \tilde{f}_{E|\mathbf{C_0}})$ *is unbiased provided that one of the following holds:*

$$
\begin{aligned}
&(a)\{\tilde{\boldsymbol{\theta}}_{\mathbf{M}}, \tilde{f}_{E|\mathbf{C_0}}\} = \{\boldsymbol{\theta}_{\boldsymbol{M}}, f_{E|\mathbf{C_0}}\}, \\
&(b)\{\tilde{B}, \tilde{\boldsymbol{\theta}}_{\mathbf{C_1}}, \tilde{f}_{E|\mathbf{C_0}}\} = \{B, \boldsymbol{\theta}_{\boldsymbol{C_1}}, f_{E|\mathbf{C_0}}\}, \text{or} \\
&(c)\{\tilde{B}, \tilde{\boldsymbol{\theta}}_{\mathbf{C_1}}, \tilde{\boldsymbol{\theta}}_{\mathbf{M}}\} = \{B, \boldsymbol{\theta}_{\mathbf{C_1}}, \boldsymbol{\theta}_{\boldsymbol{M}}\}.
\end{aligned}
$$

Proof.

$\mathbb{E}V^{eff}(\beta_0,\tilde{B}, \tilde{\boldsymbol{\theta}}_{\mathbf{M}}, \tilde{\boldsymbol{\theta}}_{\mathbf{C_1}}, \tilde{f}_{E|\mathbf{C_0}}) =$

$$
\begin{aligned}
&\mathbb{E}\Bigg[\int\limits_{m,\mathbf{c_1}} \frac{\tilde{M}^{ratio}}{\tilde{f}(e'|\mathbf{C_0})}\left\{B(m,\mathbf{c_1},e',\mathbf{C_0})-\tilde{B}(m,\mathbf{c_1},e',\mathbf{C_0})\right\}\\
&\qquad\times f(m|\mathbf{c_1},e',\mathbf{C_0})f(\mathbf{c_1}|e',\mathbf{C_0})f(e'|\mathbf{C_0})d\nu(m,\mathbf{c_1})\\
&\qquad+\int\limits_{\mathbf{c_1}}\frac{1}{\tilde{C}_1^{ratio}\tilde{f}(e|\mathbf{C_0})}\left\{\mathbb{E}\left[\tilde{B}(M,\mathbf{c_1},e',\mathbf{C_0})|\mathbf{c_1},e,\mathbf{C_0}\right]-\tilde{\mathbb{E}}\left[\tilde{B}(M,\mathbf{c_1},e',\mathbf{C_0})|\mathbf{c_1},e,\mathbf{C_0}\right]\right\}\\
&\qquad\times f(\mathbf{c_1}|e,\mathbf{C_0})f(e|\mathbf{C_0})d\nu(\mathbf{c_1})\\
&\qquad+\frac{f(e'|\mathbf{C_0})}{\tilde{f}(e'|\mathbf{C_0})}\left\{\mathbb{E}\left[\tilde{\mathbb{E}}\left[\tilde{B}(M,\mathbf{C_1},e',\mathbf{C_0})|\mathbf{C_1},e,\mathbf{C_0}\right]|e',\mathbf{C_0}\right]\right.\\
&\qquad\left.-\tilde{\mathbb{E}}\left[\tilde{\mathbb{E}}\left[\tilde{B}(M,\mathbf{C_1},e',\mathbf{C_0})|\mathbf{C_1},e,\mathbf{C_0}\right]|e',\mathbf{C_0}\right]\right\}\\
&\qquad+\tilde{\mathbb{E}}\left[\tilde{\mathbb{E}}\left[\tilde{B}(M,\mathbf{C_1},e',\mathbf{C_0})|\mathbf{C_1},e,\mathbf{C_0}\right]|e',\mathbf{C_0}\right]\\
&\qquad-\mathbb{E}\left[\mathbb{E}\left[B(M,\mathbf{C_1},e',\mathbf{C_0})|\mathbf{C_1},e,\mathbf{C_0}\right]|e',\mathbf{C_0}\right]\Bigg]
\end{aligned}
$$

Substituting under (a):

$$
\begin{aligned}
&\mathbb{E}V^{eff}(\beta_0,\tilde{B},\tilde{\boldsymbol{\theta}}_{\mathbf{M}},\tilde{\boldsymbol{\theta}}_{\mathbf{C_1}},\tilde{f}_{E|\mathbf{C_0}})=\\
&\quad\mathbb{E}\Bigg[\int\limits_{m,\mathbf{c_1}}\left\{B(m,\mathbf{c_1},e',\mathbf{C_0})-\tilde{B}(m,\mathbf{c_1},e',\mathbf{C_0})\right\}f(m|\mathbf{c_1},e,\mathbf{C_0})f(\mathbf{c_1}|e',\mathbf{C_0})d\nu(m,\mathbf{c_1})\\
&\qquad+\left\{\mathbb{E}\left[\mathbb{E}\left[\tilde{B}(M,\mathbf{C_1},e',\mathbf{C_0})|\mathbf{C_1},e,\mathbf{C_0}\right]|e',\mathbf{C_0}\right]\right.\\
&\qquad\quad\left.-\tilde{\mathbb{E}}\left[\mathbb{E}\left[\tilde{B}(M,\mathbf{C_1},e',\mathbf{C_0})|\mathbf{C_1},e,\mathbf{C_0}\right]|e',\mathbf{C_0}\right]\right\}\\
&\qquad+\tilde{\mathbb{E}}\left[\mathbb{E}\left[\tilde{B}(M,\mathbf{C_1},e',\mathbf{C_0})|\mathbf{C_1},e,\mathbf{C_0}\right]|e',\mathbf{C_0}\right]\\
&\qquad-\mathbb{E}\left[\mathbb{E}\left[B(M,\mathbf{C_1},e',\mathbf{C_0})|\mathbf{C_1},e,\mathbf{C_0}\right]|e',\mathbf{C_0}\right]\Bigg]\\
&\qquad=0
\end{aligned}
$$

Substituting under (b):

$$
\begin{aligned}
&\mathbb{E}V^{eff}(\beta_0,\tilde{B},\tilde{\boldsymbol{\theta}}_{\mathbf{M}},\tilde{\boldsymbol{\theta}}_{\mathbf{C_1}},\tilde{f}_{E|\mathbf{C_0}})=\\
&\qquad\int\limits_{\mathbf{c_1}}\left\{\mathbb{E}\left[B(M,\mathbf{c_1},e',\mathbf{C_0})|\mathbf{c_1},e,\mathbf{C_0}\right]-\tilde{\mathbb{E}}\left[B(M,\mathbf{c_1},e',\mathbf{C_0})|\mathbf{c_1},e,\mathbf{C_0}\right]\right\}f(\mathbf{c_1}|e',\mathbf{C_0})d\nu(\mathbf{c_1})\\
&\qquad+\mathbb{E}\left[\tilde{\mathbb{E}}\left[B(M,\mathbf{C_1},e',\mathbf{C_0})|\mathbf{C_1},e,\mathbf{C_0}\right]|e',\mathbf{C_0}\right]
\end{aligned}
$$

$$\left. -\mathbb{E}\left[\mathbb{E}\left[B(M,\mathbf{C_1},e',\mathbf{C_0})|\mathbf{C_1},e,\mathbf{C_0}\right]|e',\mathbf{C_0}\right] \right]$$

$$= 0$$

Substituting under (c):
$\mathbb{E}V^{eff}(\beta_0,\tilde{B},\tilde{\boldsymbol{\theta}}_{\mathbf{M}},\tilde{\boldsymbol{\theta}}_{\mathbf{C_1}},\tilde{f}_{E|\mathbf{C_0}}) = 0$, trivially. □

Thus, $\hat{\beta}_{mr}$ can be shown to be asymptotically normal under each of these scenarios using a Taylor expansion of $\mathbb{P}_n V^{eff}(\hat{\beta}_{mr},\hat{B},\hat{\boldsymbol{\theta}}_{\mathbf{M}},\hat{\boldsymbol{\theta}}_{\mathbf{C_1}},\hat{f}_{E|\mathbf{C_0}})$ and applying the central limit theorem to $n^{-1/2}\sum_i V_i^{eff}(\beta_0,B^*,\boldsymbol{\theta}_{\boldsymbol{M}}^*,\boldsymbol{\theta}_{\mathbf{C_1}}^*,f_{E|\mathbf{C_0}}^*)$.

Appendix B

Additional stabilization technique for the multiply-robust estimator

This technique is an adaptation of the approach presented by Robins et al. (2007). The idea is to carefully select regression models and an estimation strategy such that the three terms in $\hat{\beta}_{mr}$ depending on weights are empirically evaluated as null, leaving the term $\hat{B}''(e', e, \mathbf{C_0})$, which does not depend on weights. This can be accomplished with the following steps. First, fit propensity score models to estimate $f_{E|\mathbf{C_0}}$, M^{ratio}, and C_1^{ratio}. Substitute these estimates into the first term of $\hat{\beta}_{mr}$, and include the result in a set of estimating equation to solve for the Y-regression-model parameters. Next, plug in all parameters estimated thus far into the second term of $\hat{\beta}_{mr}$, and once again use the result in a set of estimating equations to solve for the M-regression-model parameters. Repeat this step with the third term of $\hat{\beta}_{mr}$ to solve for the $\mathbf{C_1}$-regression-model parameters. Finally, plugging all of these parameter estimates into $\hat{\beta}_{mr}$ leaves $\hat{B}''(e', e, \mathbf{C_0})$, as desired. If Y, M, and $\mathbf{C_1}$ are all scalar, continuous random variables, this procedure is equivalent to repeatedly fitting regression models with intercepts using weighted least squares with appropriately-chosen weights.

Appendix C

Plots comparing estimators in the PEPFAR Nigeria study

Figure C.1: $\mathcal{P}_{EMY}$-specific effects on virologic failure. The plot in each cell represents estimates for the effect with comparison-level treatment, e, equal to the first index of the cell and baseline-level treatment, e' equal to the second index of the cell. That is, comparison level treatment varies across rows and baseline level treatment varies across columns. Within each plot, the dots and vertical bars represent point estimates using the four estimators and their corresponding bootstrap confidence intervals.

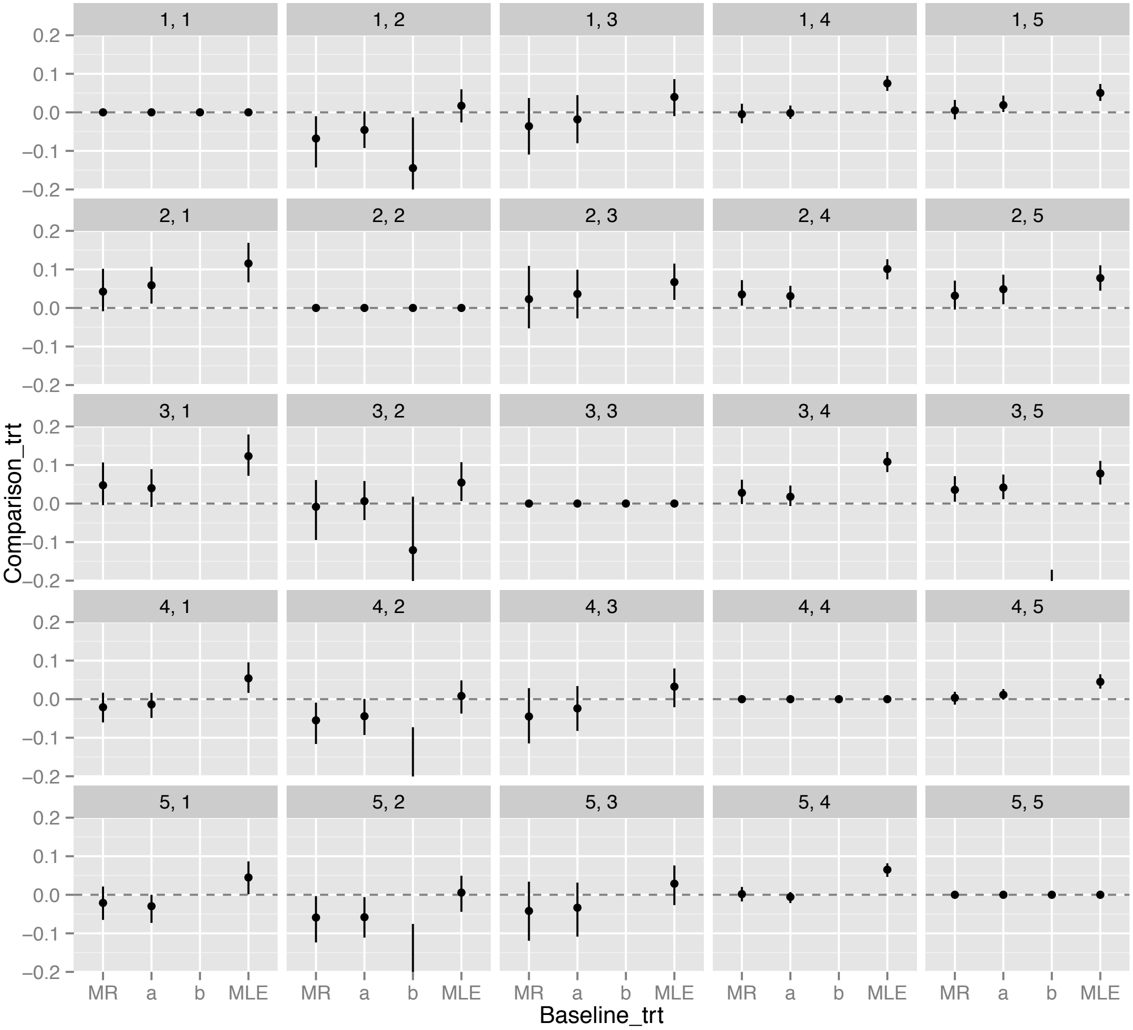


Figure C.1: (Continued)

Figure C.2: $\mathcal{P}_{EMY}$-specific effects on CD4 count. The plot in each cell represents estimates for the effect with comparison-level treatment, e, equal to the first index of the cell and baseline-level treatment, e' equal to the second index of the cell. That is, comparison level treatment varies across rows and baseline level treatment varies across columns. Within each plot, the dots and vertical bars represent point estimates using the four estimators and their corresponding bootstrap confidence intervals.

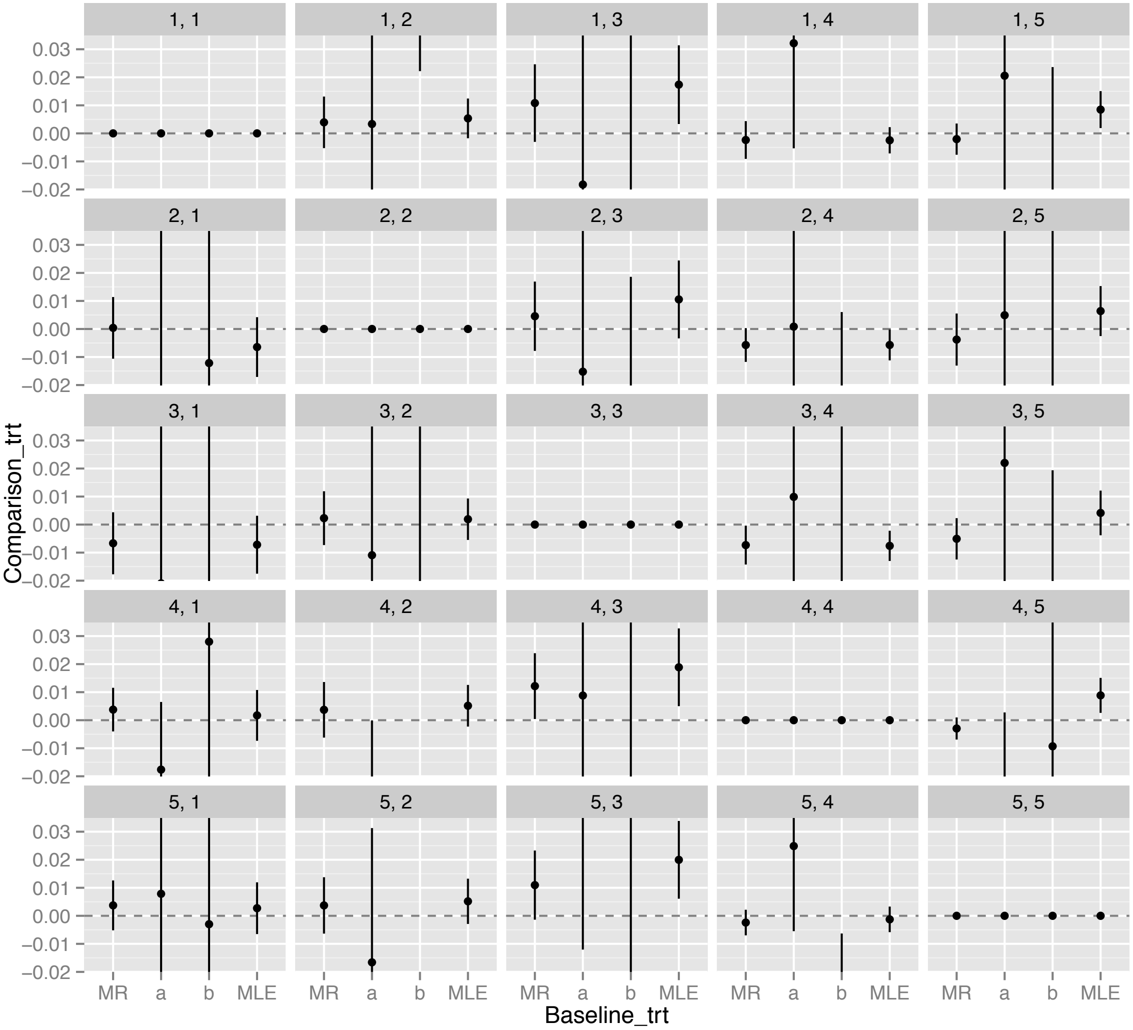


Figure C.2: (Continued)

Appendix D

Proofs for Chapter 2

Proof of Theorem 2.1. For each level m and y, define $\pi_1(m,y) = \mathrm{pr}\{Y(a,m) = y\}$ and $\pi_2(m) = \mathrm{pr}\{M(a^*) = m\}$. There exist $U_1(m,y), U_2(m) \sim \mathcal{U}(0,1)$ such that $I\{Y(a,m) = y\} = I\{U_1(m,y) \leq \pi_1(m,y)\}$ and $I\{M(a^*) = m\} = I\{U_2(m) \leq \pi_2(m)\}$. The joint distribution $F_{U_1(m,y),U_2(m)}$, then, is a bivariate copula, for which Fréchet–Hoeffding sharp bounds exist. Applying these to $\mathrm{pr}\left\{Y(a,m) = y, M(a^*) = m\right\} = F_{U_1(m,y),U_2(m)}\left\{\pi_1(m,y), \pi_2(m)\right\}$, we have

$$\begin{gathered}\max\left[0, \mathrm{pr}\{M(a^*) = m\} + \mathrm{pr}\{Y(a,m) = y\} - 1\right] \\ \leq \mathrm{pr}\left\{Y(a,m) = y, M(a^*) = m\right\} \leq \\ \min\left[\mathrm{pr}\{M(a^*) = m\}, \mathrm{pr}\{Y(a,m) = y\}\right].\end{gathered}$$

Applying these bounds to each summand in

$$E[Y\{a, M(a^*)\}] = \sum_{m,y} y\mathrm{pr}\{Y(a,m) = y, M(a^*) = m\}$$

yields the result. □

Proof of Theorem 2.3. Let $\pi_{r,r^*} = \mathrm{pr}\left\{R(a) = r, R(a^*) = r^*\right\}$, π be the vectorization of the matrix $[\pi_{r,r^*}]$, and δ be the vectorization of the matrix $[\pi_{r,r^*}]_{-p,-p}$, i.e., the vectorization of the matrix π with row p and column p removed. Equation (2.1) can now be expressed as $\gamma_0 = x^T\pi$, which is identified in x, but not π. Conditional on the marginal probabilities, which are identified, the joint probabilities have $(p-1)^2$ degrees of freedom, and can be expressed as $\pi = B\delta + d$. Since $x^T B\delta$ is linear in δ and each element of δ is constrained by

$$\max\left\{0, \mathrm{pr}(R = r \mid A = a) + \mathrm{pr}(R = r^* \mid A = a^*) - 1\right\}$$

$$\leq \pi_{r,r^*} \leq$$

$$\min\left\{\mathrm{pr}(R = r \mid A = a), \mathrm{pr}(R = r^* \mid A = a^*)\right\},$$

the proposed linear programming problem will yield the δ that optimizes $x^T B\delta$, and hence $x^T(B\delta + d)$. Thus, γ_0 will be bounded by $x^T(B\delta + d)$ evaluated at the minimizing and maximizing linear programming solutions δ_L and δ_U. □

Appendix E

Proofs and additional figures for Chapter 3

Proof of Theorem 3.1. Let $\bar{\alpha}$ be the true value of α. Under $\mathcal{M}_A$,

$$
\begin{aligned}
E\{U(\bar{\alpha})\} =& E[\{\ell(C)Y + m(C)\}(A - E(A \mid C, X^*) - \bar{\alpha}_2^T\varepsilon^*)] \\
=& E[\{\ell(C)Y + m(C)\}A] - E[E\{\ell(C)Y + m(C) \mid C, X^*\}E(A \mid C, X^*)] \\
& - \alpha_2^T E[\varepsilon^*]E[\{\ell(C)Y + m(C)\}] \\
=& E[\{\ell(C)Y + m(C)\}A] - E\left\{E\left(E[\{\ell(C)Y + m(C)\}A \mid A, C, X^*] \mid C, X^*\right)\right\} \\
=& 0,
\end{aligned}
$$

since $Y \perp\!\!\!\perp \epsilon^*$, $C \perp\!\!\!\perp \epsilon^*$, and $E(Y \mid A, C, X^*) = E(Y \mid C, X^*)$, which is implied by H_0 and Assumption 1. The regularity conditions on $U(\alpha)$ are sufficient to ensure that $\bar{\alpha}$ is a local minimum of $n\hat{U}_n(\alpha)^T\hat{\Omega}_n^{-1}\hat{U}_n(\alpha)$, and hence α is locally identified under H_0. Then because $\dim[U(\alpha)] = \dim(\alpha) + q$, $E\{U(\alpha)\} = 0$ is an overidentified moment restriction, and the statistic $\chi^2_{\mathrm{robust}_A}$ has a limiting distribution of χ^2_q. □

Proof of Theorem 3.2. Let $\bar{\gamma}$ and $\bar{\alpha}_2$ be the true values of γ and α_2 respectively. Under $\mathcal{M}_Y$, $E\{S(\bar{\gamma})\} = 0$ and

$$
\begin{aligned}
& E\left[k(C)\left\{Y - g_Y(C;\bar{\gamma})\right\}\left\{A - \bar{\alpha}_2^T X\right\}\right] \\
=& E\left[k(C)\left\{Y - E\left(Y \mid C\right)\right\}\left\{A - E(A \mid C, X^*) + E(A \mid C, X^* = 0) - \bar{\alpha}_2^T\varepsilon^*\right\}\right] \\
=& E\left[k(C)\left\{Y - E\left(Y \mid C\right)\right\}A\right] - E\left[k(C)E\left\{Y - E\left(Y \mid C\right) \mid C, X^*\right\}E(A \mid C, X^*)\right] \\
& + E\left[k(C)E\left\{Y - E(Y \mid C) \mid C\right\}\left\{E(A \mid C, X^* = 0) - \bar{\alpha}_2^T E(\varepsilon^*)\right\}\right]
\end{aligned}
$$

$$
\begin{aligned}
&= E\left[k(C)\left\{Y - E\left(Y \mid C\right)\right\}A\right] - E\left\{k(C)E\left(E[\{Y - E\left(Y \mid C\right)\}A \mid A, C, X^*] \mid C, X^*\right)\right\}\\
&= 0,
\end{aligned}
$$

since $Y \perp\!\!\!\perp \epsilon^*$, $C \perp\!\!\!\perp \epsilon^*$, and $E(Y \mid A, C, X^*) = E(Y \mid C, X^*)$, which is implied by H_0 and Assumption 1. The regularity conditions on $U(\alpha, \gamma)$ are sufficient to ensure that $(\bar{\alpha}_2, \bar{\gamma})$ is a local minimum of $n\hat{U}_n(\alpha, \gamma)^T \hat{\Omega}_n^{-1} \hat{U}_n(\alpha, \gamma)$, and hence (α_2, γ) is locally identified under H_0. Then because $\dim[U(\alpha, \gamma)] = \dim(\alpha_2) + \dim(\gamma) + q$, $E\{U(\alpha, \gamma)\} = 0$ is an overidentified moment restriction, and the statistic $\chi^2_{\text{robust}_Y}$ has a limiting distribution of χ^2_q.

□

Proof of Theorem 3.3. Let $\bar{\gamma}$ be the true value of γ under $\mathcal{M}_Y$ and $\bar{\alpha}$ be the true value of α under $\mathcal{M}_A$. Under $\mathcal{M}_A$, there exists some $\tilde{\gamma}$ such that $E\{S(\tilde{\gamma})\} = 0$. That

$$
E\left[\begin{array}{c} k(C)\Delta_Y(\tilde{\gamma})\Delta_A(\bar{\alpha}) \\ \{\ell(C)Y + m(C)\}\,\Delta_A(\bar{\alpha}) \end{array}\right] = 0
$$

follows from the unbiasedness of $U(\alpha)$ shown in the proof of Theorem 3.1.

Under $\mathcal{M}_Y$, $E\{S(\bar{\gamma})\} = 0$ and for any α_1,

$$
\begin{aligned}
&E\left[k(C)\left\{Y - g_Y(C;\bar{\gamma})\right\}\left\{A - g_A(C;\alpha_1) - \bar{\alpha}_2^T X\right\}\right]\\
=&E\left[k(C)\left\{Y - E\left(Y \mid C\right)\right\}\left\{A - E(A \mid C, X^*) + E(A \mid C, X^* = 0) - g_A(C;\alpha_1) - \bar{\alpha}_2^T \varepsilon^*\right\}\right]\\
=&E\left[k(C)\left\{Y - E\left(Y \mid C\right)\right\}A\right] - E\left[k(C)E\left\{Y - E\left(Y \mid C\right) \mid C, X^*\right\}E(A \mid C, X^*)\right]\\
&+ E\left[k(C)E\left\{Y - E(Y \mid C) \mid C\right\}\left\{E(A \mid C, X^* = 0) - g_A(C;\alpha_1) - \bar{\alpha}_2^T E(\varepsilon^*)\right\}\right]\\
=&E\left[k(C)\left\{Y - E\left(Y \mid C\right)\right\}A\right] - E\left\{k(C)E\left(E[\{Y - E\left(Y \mid C\right)\}A \mid A, C, X^*] \mid C, X^*\right)\right\}\\
=&0,
\end{aligned}
$$

since $Y \perp\!\!\!\perp \epsilon^*$, $C \perp\!\!\!\perp \epsilon^*$, and $E(Y \mid A, C, X^*) = E(Y \mid C, X^*)$, which is implied by H_0 and Assumption 1. Finally, there exists some $\tilde{\alpha}_1$ such that

$$
E\left[\{\ell(C)Y + m(C)\}\,\Delta_A(\tilde{\alpha}_1, \bar{\alpha}_2)\right] = 0.
$$

Thus, under any law in $\mathcal{M}_\cup$, $E\{U(\alpha, \gamma)\} = 0$ has a solution under H_0. The regularity conditions on $U(\alpha, \gamma)$ are sufficient to ensure that $(\bar{\alpha}, \bar{\gamma})$ is a local minimum of $n\hat{U}_n(\alpha, \gamma)^T \hat{\Omega}_n^{-1} \hat{U}_n(\alpha, \gamma)$, and hence (α, γ) is locally identified under H_0. Then because $\dim[U(\alpha, \gamma)] = \dim(\alpha) + \dim(\gamma) + q$,

$E\{U(\alpha,\gamma)\}=0$ is an overidentified moment restriction, and the statistic χ^2_{dr} has a limiting distribution of χ^2_q.

□

Proof of Theorem 3.4. First, reparameterize the propensity-score model as

$$\text{logit}\,\text{Pr}(A=1\mid C,X^*)=\alpha_0+g(C;\alpha_1)+\alpha_2^TX^*,$$

for binary A or

$$\text{log}E(A\mid C,X^*)=\alpha_0+g(C;\alpha_1)+\alpha_2^TX^*,$$

for count A, where $g(0;\alpha_1)=0$, such that α_0 is a scalar intercept and α_1 has dimension p_1-1. When A is binary, for the true value $\bar\alpha$ of α, we have

$$\begin{aligned}
& E\left\{\{\ell(C)Y+m(C)\}\exp\left(-\bar\alpha_2^TXA\right)\left[A-\text{expit}\left(\alpha_0^*+g(C;\bar\alpha_1)\right)\right]\right\}\\
= \;& E\left\{\{\ell(C)Y+m(C)\}\exp\left(-\bar\alpha_2^TX^*A-\bar\alpha_2^T\varepsilon^*A\right)\left[A-\text{expit}\left(\alpha_0^*+g(C;\bar\alpha_1)\right)\right]\right\}\\
= \;& E\left\{\{\ell(C)Y+m(C)\}\exp\left(-\bar\alpha_2^TX^*A\right)E\left(\exp\left(-\bar\alpha_2^T\varepsilon^*A\right)|A,Y,C,X^*\right)\right.\\
& \left.\times\left[A-\text{expit}\left(\alpha_0^*+g(C;\bar\alpha_1)\right)\right]\right\}\\
= \;& E\left\{\{\ell(C)Y+m(C)\}\exp\left(-\bar\alpha_2^TX^*A\right)\exp\left(KA\right)\left[A-\text{expit}\left(\alpha_0^*+g(C;\bar\alpha_1)\right)\right]\right\}
\end{aligned}$$

where $\exp(K)=E\left(\exp\left(-\bar\alpha_2^T\varepsilon^*\right)\right)$ is the moment generating function of ε^* evaluated at $-\bar\alpha_2$. We then note that the joint density of (A,X^*) given C can be expressed as

$$f\left(A,X^*|C\right)=\frac{f\left(X^*|A=0,C\right)\exp\left(\bar\alpha_2^TX^*A\right)f\left(A|X^*=0,C\right)}{t(C)}$$

where $t(C)$ is a normalizing constant. We then have that

$$\begin{aligned}
& E\left\{\{\ell(C)Y+m(C)\}\exp\left(-\bar\alpha_2^TX^*A\right)\exp\left(KA\right)\left[A-\text{expit}\left(\alpha_0^*+g(C;\bar\alpha_1)\right)\right]\right\}\\
= \;& E\int_x\sum_a\frac{f\left(x|A=0,C\right)\exp\left(\bar\alpha_2^Txa\right)f\left(a|X=0,C\right)}{t(C)}\\
& \times\left[\ell(C)E\left\{Y\mid a,x,C\right\}+m(C)\right]\exp\left(-\bar\alpha_2^Txa\right)\exp\left(Ka\right)\left[a-\text{expit}\left(\alpha_0^*+g(C;\bar\alpha_1)\right)\right]dx\\
= \;& E\int_x\left[\ell(C)E\left\{Y\mid x,C\right\}+m(C)\right]f\left(x|A=0,C\right)t(C)^{-1}dx\\
& \times\sum_a f\left(a|X=0,C\right)\exp\left(Ka\right)\left[a-\text{expit}\left(\alpha_0^*+g(C;\bar\alpha_1)\right)\right]
\end{aligned}$$

$$
\begin{aligned}
&= E\int_x \left[\ell(C)E\left\{Y \mid x, C\right\} + m(C)\right] f\left(x|A=0,C\right) t(C)^{-1}dx \frac{1+\exp\left(\alpha_0^* + g(C;\bar{\alpha}_1)\right)}{1+\exp\left(\bar{\alpha}_0 + g(C;\bar{\alpha}_1)\right)} \\
&\quad \times \sum_a \frac{\exp\left(\alpha_0^* a + g(C;\bar{\alpha}_1)a\right)}{1+\exp\left(\alpha_0^* + g(C;\bar{\alpha}_1)\right)}\left[a - \text{expit}\left(\alpha_0^* + g(C;\bar{\alpha}_1)\right)\right] \\
&= 0
\end{aligned}
$$

where $\alpha_0^* = K + \bar{\alpha}_0$.

When A is a count, let $\exp(K) = E\left\{\exp(\bar{\alpha}_2^T\varepsilon^*)\right\}$, i.e., the moment generating function of ε^* evaluated at $\bar{\alpha}_2$. For the true value $\bar{\alpha}$ of α, we have

$$
\begin{aligned}
&E\left(Y\left[A - \exp\left\{\alpha_0^* + g(C;\bar{\alpha}_1) + \bar{\alpha}_2^T X\right\}\right]\right) \\
=&E(YA) - E\left[Y\exp\left\{\alpha_0^* + g(C;\bar{\alpha}_1) + \bar{\alpha}_2^T X^*\right\} E\left\{\exp(\bar{\alpha}_2^T\varepsilon^*)\right\}\right] \\
=&E(YA) - E\left[Y\exp\left\{\bar{\alpha}_0 + g(C;\bar{\alpha}_1) + \bar{\alpha}_2^T X^*\right\}\right] \\
=&E(YA) - E\left\{E(Y \mid C, X^*)E(A \mid C, X^*)\right\} \\
=&E(YA) - E\left[E\left\{E(YA \mid A, C, X^*) \mid C, X^*\right\}\right] \\
=&0,
\end{aligned}
$$

where $\alpha_0^* = K + \bar{\alpha}_0$.

Thus, in either case, $E\{U(\alpha_0^*, \bar{\alpha}_1, \bar{\alpha}_2)\} = 0$ under H_0. The regularity conditions on $U(\alpha)$ are sufficient to ensure that $(\alpha_0^*, \bar{\alpha}_1, \bar{\alpha}_2)$ is a local minimum of $n\hat{U}_n(\alpha)^T\hat{\Omega}_n^{-1}\hat{U}_n(\alpha)$, and hence α is locally identified under H_0. Then because $\dim[U(\alpha)] = \dim(\alpha) + q$, $E\{U(\alpha)\} = 0$ is an overidentified moment restriction, and the statistic $\chi^2_{\text{robust}_A}$ has a limiting distribution of χ^2_q. □

Theorem E.1. *Let $\hat{\beta}(\ell, m)$ be the estimator solving $\mathbb{P}_n U(\ell, m; \beta) = 0$ corresponding to the moment functions in Theorem 3.1, and*

$$
d(C) \equiv E\{\Delta(\alpha)^2 H(\psi)^2 \mid C\}E\{\Delta(\alpha)^2 \mid C\} - E\{\Delta(\alpha)^2 H(\psi) \mid C\}E\{\Delta(\alpha)^2 H(\psi) \mid C\}.
$$

For

$$
\ell^*(C) = d(C)^{-1}\begin{bmatrix} E\left\{\Delta(\alpha)^2 \mid C\right\}E\left\{\Delta(\alpha)A \mid C\right\} \\ -\mathrm{Cov}\{\Delta(\alpha)^2, H(\psi) \mid C\} \\ E\left\{\Delta(\alpha)^2 \mid C\right\}E\left\{\Delta(\alpha)A \mid C\right\}\left\{\nabla_{\alpha_C} g(C;\alpha_C)\right\}^T \\ E(H(\psi)X \mid C)E\left\{\Delta(\alpha)^2 \mid C\right\} - E(X \mid C)E\left\{\Delta(\alpha)^2 H(\psi) \mid C\right\} \end{bmatrix}
$$

and

$$m^*(C) = d(C)^{-1}\left[\begin{array}{c} -E\left\{\Delta(\alpha)^2 H(\psi) \mid C\right\} E\left\{\Delta(\alpha) A \mid C\right\} \\ -\mathrm{Cov}\{\Delta(\alpha)^2 H(\psi), H(\psi) \mid C\} \\ -E\left\{\Delta(\alpha)^2 H(\psi) \mid C\right\} E\left\{\Delta(\alpha) A \mid C\right\}\left\{\nabla_{\alpha_C} g(C;\alpha_C)\right\}^T \\ E(H(\psi)X \mid C) E\left\{\Delta(\alpha)^2 H(\psi) \mid C\right\} - E(X \mid C) E\left\{\Delta(\alpha)^2 H(\psi)^2 \mid C\right\} \end{array}\right],$$

$\hat{\beta}(\ell^*, m^*)$ *achieves the minimum asymptotic variance of all estimators in the class of estimators defined by these estimating equations. The corresponding variance is* $\mathbb{E}\left[\mathbf{U}(\ell^*, m^*; \beta)\mathbf{U}(\ell^*, m^*; \beta)^T\right]$.

Proof. By Theorem 5.3 in Newey and McFadden (1994), if an optimal estimator $\hat{\beta}(\tilde{\ell}, \tilde{m})$ exists within the class $\{\hat{\beta}(\ell, m): \ \ell \in \mathcal{L}, m \in \mathcal{M}\}$, the functions $\tilde{\ell}$ and $\tilde{m}$ are guaranteed to satisfy

$$-\mathbb{E}\left[\frac{\partial}{\partial\beta}\mathbf{U}(\ell, m; \beta)\right] = \mathbb{E}\left[\mathbf{U}(\ell, m; \beta)\mathbf{U}(\tilde{\ell}, \tilde{m}; \beta)^T\right] \tag{E.1}$$

for all functions ℓ and m, and the estimator will have variance equal to $\mathbb{E}\left[\mathbf{U}(\tilde{\ell}, \tilde{m}; \beta)\mathbf{U}(\tilde{\ell}, \tilde{m}; \beta)^T\right]$. Thus it suffices to show that $\ell^*(C)$ and $m^*(C)$ satisfy (E.1). We have

$$\begin{aligned} -E\left[\frac{\partial}{\partial\beta}\mathbf{U}(\ell, m; \beta)\right] &= E\left[\ell(C)A\Delta(\alpha), \ \ \ell(C)H(\psi) + m(C),\right. \\ &\quad \left.\{\ell(C)H(\psi) + m(C)\}\{\nabla_{\alpha_C} g(C;\alpha_C)\}^T, \ \ \{\ell(C)H(\psi) + m(C)\}X^T\right] \\ &= E\left\{[\ell(C), m(C)]\left[\begin{array}{c} V \\ W \end{array}\right]\right\}, \end{aligned}$$

where $V = [V_1, V_2, V_3, V_4] \equiv [A\Delta(\alpha), H(\psi), H(\psi)\{\nabla_{\alpha_C} g(C;\alpha_C)\}^T, H(\psi)X^T]$ and $W = [W_1, W_2, W_3, W_4] \equiv [0, 1, \{\nabla_{\alpha_C} g(C;\alpha_C)\}^T, X^T]$. If we partition the components of the functions (ℓ^*, m^*) into (ℓ_1^*, m_1^*), (ℓ_2^*, m_2^*), (ℓ_3^*, m_3^*), and (ℓ_4^*, m_4^*), where ℓ_1^*, m_1^*, ℓ_2^*, and m_2^* are scalar functions, ℓ_3^* and m_3^* are p_1 dimensional, and ℓ_4^* and m_4^* are p_2 dimensional, then

$$\begin{aligned} &\mathbb{E}\left[\mathbf{U}(\ell, m; \beta)\mathbf{U}(\ell^*, m^*; \beta)^T\right] = E\left[\Delta(\alpha)^2\{\ell(C)H(\psi) + m(C)\}\{\ell^*(C)H(\psi) + m^*(C)\}^T\right] \\ =&E\left[\Delta(\alpha)^2\{\ell(C)H(\psi) + m(C)\}\left[\ell_1^*(C)H(\psi) + m_1^*(C), \ \ \{\ell_2^*(C)H(\psi) + m_2^*(C)\}^T,\right.\right. \\ &\quad \left.\left.\{\ell_3^*(C)H(\psi) + m_3^*(C)\}^T, \ \ \ell_4^*(C)H(\psi) + m_4^*(C)\right]\right]. \end{aligned}$$

Thus, we can solve (E.1) by partitioning it into four independent equations corresponding to the partition of ℓ^* and m^*, i.e., for $k = 1, 2, 3, 4$,

$$E\left[\Delta(\alpha)^2\{\ell(C)H(\psi) + m(C)\}\left\{\ell_k^*(C)H(\psi) + m_k^*(C)\right\}\right] = E\left\{\ell(C)V_k + m(C)W_k\right\}.$$

$$
\begin{aligned}
&E\left[E\left[\Delta(\alpha)^2\left\{\ell_k^*(C)H(\psi)+m_k^*(C)\right\}H(\psi)\mid C\right]\ell(C)\right.\\
&\qquad\left.-E\left[\Delta(\alpha)^2\left\{\ell_k^*(C)H(\psi)+m_k^*(C)\right\}\mid C\right]m(C)\right]=E\left\{\ell(C)V_k+m(C)W_k\right\}
\end{aligned}
$$

$\Leftrightarrow$

$$
\begin{aligned}
&E\left[E\left[\Delta(\alpha)^2\left\{\ell_k^*(C)H(\psi)+m_k^*(C)\right\}H(\psi)-V_k\mid C\right]\ell(C)\right.\\
&\qquad\left.-E\left[\Delta(\alpha)^2\left\{\ell_k^*(C)H(\psi)+m_k^*(C)\right\}-W_k\mid C\right]m(C)\right]=0
\end{aligned}
$$

$\Leftrightarrow$

$$
\begin{aligned}
E[\Delta(\alpha)^2H(\psi)^2\mid C]\ell_k^*(C)+E[\Delta(\alpha)^2H(\psi)\mid C]m_k^*(C)-E(V_k\mid C)&=0\\
E[\Delta(\alpha)^2H(\psi)\mid C]\ell_k^*(C)+E[\Delta(\alpha)^2\mid C]m_k^*(C)-E(W_k\mid C)&=0
\end{aligned}
$$

$\Leftrightarrow$

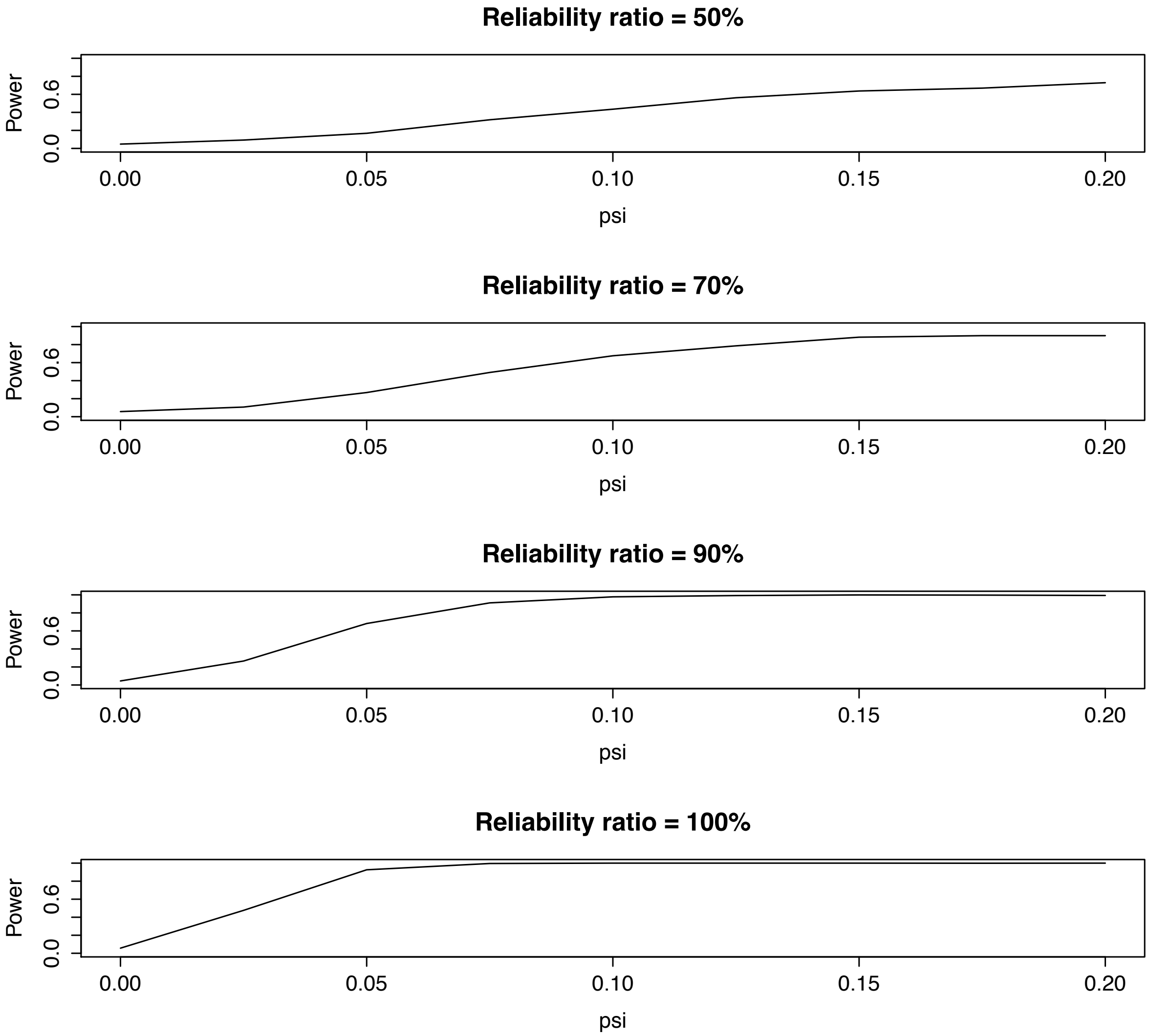
$$
\left[\begin{array}{c}\ell_k^*(C)\\ m_k^*(C)\end{array}\right]=\left[\begin{array}{cc}E\{\Delta(\alpha)^2H(\psi)^2\mid C\} & E\{\Delta(\alpha)^2H(\psi)\mid C\}\\ E\{\Delta(\alpha)^2H(\psi)\mid C\} & E\{\Delta(\alpha)^2\mid C\}\end{array}\right]^{-1}\left[\begin{array}{c}E(V_k\mid C)\\ E(W_k\mid C)\end{array}\right]
$$

given that Pr$\{d(C)\neq 0\}=1$ The second implication can be seen to hold by recognizing the necessity of the first equation when $\ell(C)=E\left[\Delta(\alpha)^2\left\{\ell_k^*(C)H(\psi)+m_k^*(C)\right\}H(\psi)-V_k\mid C\right]$ and $b(C)=0$ and the necessity of the second equation when $\ell(C)=0$ and $b(C)=E\left[\Delta(\alpha)^2\left\{\ell_k^*(C)H(\psi)+m_k^*(C)\right\}-W_k\mid C\right]$. Thus, (E.1) is solved by

$$
\begin{aligned}
\ell_1^*(C)&=E\left\{\Delta(\alpha)^2\mid C\right\}E\left\{\Delta(\alpha)A\mid C\right\}d(C)^{-1}\\
m_1^*(C)&=-E\left\{\Delta(\alpha)^2H(\psi)\mid C\right\}E\left\{\Delta(\alpha)A\mid C\right\}d(C)^{-1}\\
\ell_2^*(C)&=-\mathrm{Cov}\{\Delta(\alpha)^2,H(\psi)\mid C\}d(C)^{-1}\\
m_2^*(C)&=-\mathrm{Cov}\{\Delta(\alpha)^2H(\psi),H(\psi)\mid C\}d(C)^{-1}\\
\ell_3^*(C)&=\ell_1^*(C)\{\nabla_{\alpha_C}g(C;\alpha_C)\}^T\\
m_3^*(C)&=m_1^*(C)\{\nabla_{\alpha_C}g(C;\alpha_C)\}^T\\
\ell_4^*(C)&=\left[E(H(\psi)X\mid C)E\left\{\Delta(\alpha)^2\mid C\right\}-E(X\mid C)E\left\{\Delta(\alpha)^2H(\psi)\mid C\right\}\right]d(C)^{-1}\\
m_4^*(C)&=\left[E(H(\psi)X\mid C)E\left\{\Delta(\alpha)^2H(\psi)\mid C\right\}-E(X\mid C)E\left\{\Delta(\alpha)^2H(\psi)^2\mid C\right\}\right]d(C)^{-1}.
\end{aligned}
$$

□

Figure E.1: Power simulation results for n=1000.


Reliability ratio = 50%

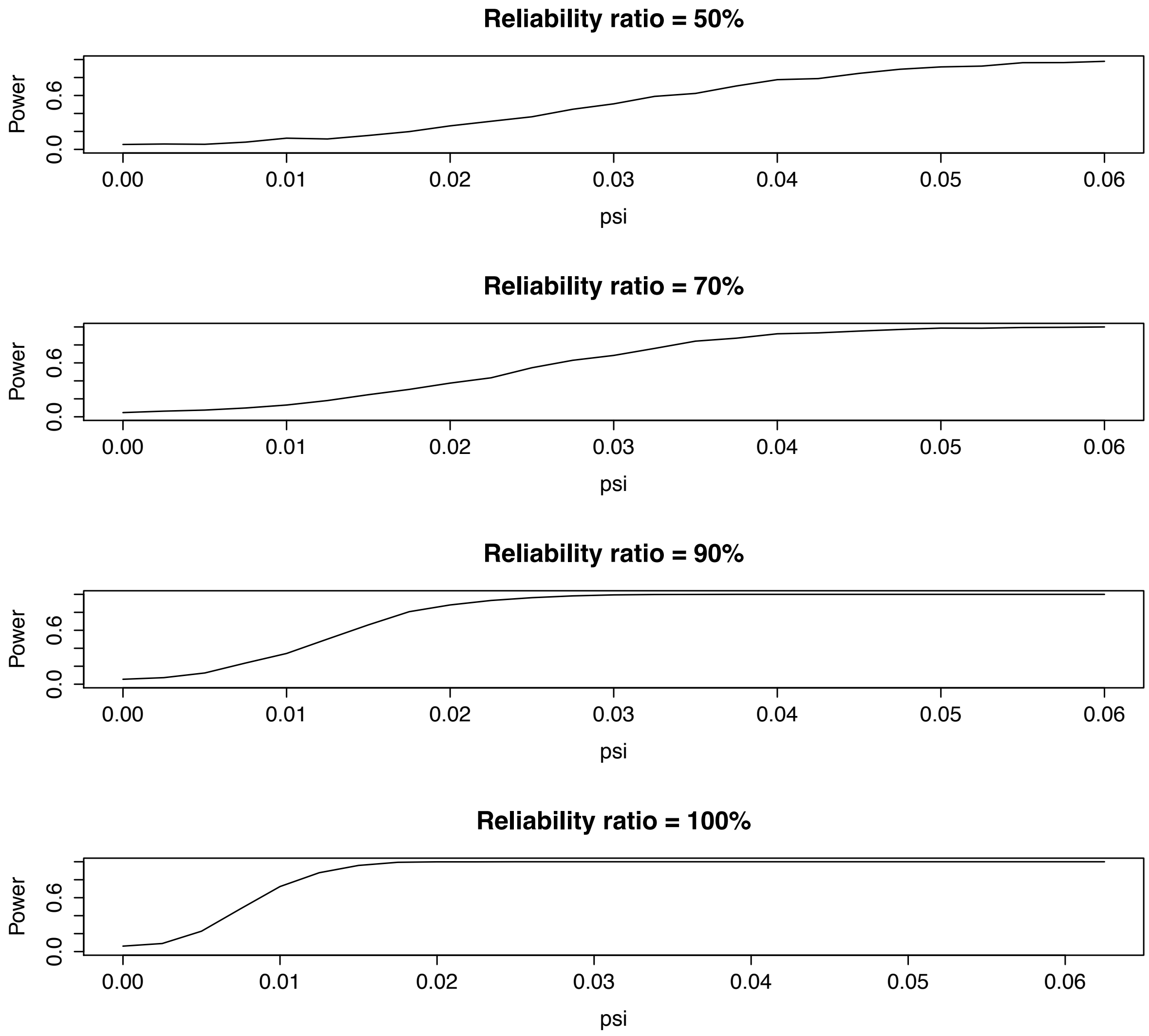
Reliability ratio = 70%

Reliability ratio = 90%

Reliability ratio = 100%

Figure E.1: (Continued)

Figure E.2: Power simulation results for n=10,000.



Reliability ratio = 50%

Reliability ratio = 70%

Reliability ratio = 90%

Reliability ratio = 100%

Figure E.2: (Continued)